

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 2, 2002, 13:32:39 ; Search time 12674 Seconds

(without alignments)  
1193.777 Million cell updates/sec

Title: US-08-973-363-1

Perfect score: 723

Sequence: 1 CCCGGTCGGAGGTTTCRAAG.....CTTGAACCTCCGACCGGG 723

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: gb\_ba:\*\*
- 2: gb\_hgt:\*\*
- 3: gb\_in:\*\*
- 4: gb\_ov:\*\*
- 5: gb\_ov:\*\*
- 6: gb\_pat:\*\*
- 7: gb\_ph:\*\*
- 8: gb\_pl:\*\*
- 9: gb\_pr:\*\*
- 10: gb\_ro:\*\*
- 11: gb\_sts:\*\*
- 12: gb\_sy:\*\*
- 13: gb\_un:\*\*
- 14: gb\_vl:\*\*
- 15: em\_ba:\*\*
- 16: em\_fun:\*\*
- 17: em\_hum:\*\*
- 18: em\_in:\*\*
- 19: em\_mu:\*\*
- 20: em\_on:\*\*
- 21: em\_or:\*\*
- 22: em\_ov:\*\*
- 23: em\_pat:\*\*
- 24: em\_ph:\*\*
- 25: em\_pl:\*\*
- 26: em\_ro:\*\*
- 27: em\_sts:\*\*
- 28: em\_un:\*\*
- 29: em\_vl:\*\*
- 30: em\_htg\_hum:\*\*
- 31: em\_htg\_inv:\*\*
- 32: em\_htg\_other:\*\*
- 33: em\_htgo\_inv:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
-----	-----	-----	-----	-----	-----

1	723	100.0	723	6	A58682
2	78	10.8	661	5	AF096788
3	62.8	8.7	145123	2	AC092403
4	62	8.6	124309	5	AC091708
5	61.2	8.5	230	5	AF076669
6	59.4	8.2	156014	5	AF003796
7	57.8	8.0	31793	5	GGU83833
8	57	7.9	187259	5	AP003795
9	55.4	7.7	109569	5	AC084760
10	55.2	7.6	156014	5	AF003796
11	55	7.6	5805	5	GA012570
12	54.8	7.6	12007	6	AX345619
13	54.6	7.6	124309	5	AC091708
14	54.2	7.5	314	5	AF111666
15	53.8	7.4	165108	2	AC092081
16	53.6	7.4	578	11	AL592916
17	53.4	7.4	34980	6	AX345531
18	52.6	7.3	2268	5	D7643352
19	52.2	7.2	187439	5	AF003795
20	51.8	7.2	1880	5	AF030300
21	51.8	7.2	3643	5	CHKNAKATPP
22	51.4	7.1	145123	2	AC092403
23	51.2	7.1	18585	6	AX281498
24	51.2	7.1	74412	2	AC020767
25	51	7.1	5526	5	CHKIGLVJC
26	50.8	7.0	13659	5	AF324491
27	50.8	7.0	14157	5	APL314752
28	50.8	7.0	15121	6	AX344830
29	50.8	7.0	48796	5	APL314754
30	50.8	7.0	165108	2	AC092081
31	50.6	7.0	6311	5	CHKCRYAA
32	50.6	7.0	135273	5	AC084761
33	50	6.9	9504	6	AX323793
34	49.8	6.9	7669	6	AX251351
35	49.8	6.9	7669	6	AX346748
36	49.8	6.9	349980	6	AX344570
37	49.6	6.9	46022	5	GA289777
38	49.6	6.9	46022	5	GA289778
39	49.6	6.9	46022	5	GA289779
40	49.6	6.9	349980	6	AX344568
41	49.4	6.8	164715	2	AC096683
42	49.2	6.8	86274	2	AC092869
43	49.2	6.8	118314	5	AC084011
44	49	6.8	117193	2	AC098851
45	48.6	6.7	152857	2	AF003457

ALIGNMENTS

RESULT	1	A58682	Sequence 1 from Patent WO9639505.	723 bp	DNA	linear	PAT 06-MAR-1998
LOCUS							
DEFINITION							
ACCESSION		A58682					
VERSION		A58682.1	GI:3714245				
KEYWORDS							
SOURCE		unidentified.					
ORGANISM		unclassified.					
REFERENCE		1 (bases 1 to 723)					
AUTHORS		Griffiths, R. and Tiwari, B.					
TITLE		AVIAN GHG GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN BIRDS					
JOURNAL		Patent: WO 9639505-A 1 12-DEC-1996;					
COMMENT		ISIS INNOVATION (GB)					
FEATURES		Other publication AU 5906996 961224.					
source		location/Qualifiers					
		1..723					
		/organism="unidentified"					
		/db_xref="taxon:32644"					
BASE COUNT		188 a 109 c 164 g 262 t.					
ORIGIN							

FEATURES	source
TITLE	Isolation and characterization of microsatellites in albatrosses
JOURNAL	Mol. Ecol. 8 (2), 338-341 (1999)
MEDLINE	99164945
REFERENCE	2 (bases 1 to 661)
AUTHORS	Burg, T. M.
TITLE	Direct Submission
JOURNAL	Submitted (05-OCT-1998) Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, UK
LOCATION/QUALIFIERS	Location/Qualifiers
1. .661	
/organism="Diomedea exulans"	
/db_xref="taxon:46550"	
1. .661	
/note="microsatellite De2"	
/rpt_type=tandem	
BASE COUNT	181 a 185 c 127 g 168 t
ORIGIN	
Query Match	10.8%; Score 78; DB 5; Length 661;
Best Local Similarity	75.3%; Pred. No. 5.le-07;
Matches 110;	Conservative 0; Mismatches 35; Indels 1; Gaps 1;
QY	17 AAGGATGACTAGATGGGACCTTAGTGCCATGGTGTGACAGGTGATGGTTGGTC 76
Db	267 AAGTATTAAAGATGTGGCACCATTAGGACATGGTTAGTGGACATGGTGTGGGTT 208
QY	77 AAAAGTGGACTCGATCTCAGACTCTTTTTCACGCTTAATAATCTATGAATCTG 136
Db	207 GATGGTGGACTTGGATCTTAGAGTCTTTTCCACCTTAATGATCTATG-ATTCTA 149
QY	137 TAATTTTATCTTGTGATCTTTTGGC 162
Db	148 TGATCTATGAAGATGATGATGTC 123
RESULT	3
AC092403/c	
LOCUS	AC092403 145123 bp DNA linear HTG 30-JAN-2002
DEFINITION	Gallus gallus clone 100N11, WORKING DRAFT SEQUENCE, 5 ordered pieces.
ACCESSION	AC092403
VERSION	AC092403.2 GI:18425382
KEYWORDS	HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE	chicken.
ORGANISM	Gallus gallus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Arcopterygia; Aves; Neognathae; Galliformes; Phasianidae;
AUTHORS	Phasianinae; Gallus.
REFERENCE	1 (bases 1 to 145123)
AUTHORS	Akhter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Bouffard, C.G., Breen, K., Brinkley, C., Brooks, S., Dietrich, N.L., Granice, S., Guan, X., Gupta, J., Heghghi, P., Ho, S.-L., Idol, J.R., Karlins, E., Laric, P., Lee-tin, S.-Q., Legaspi, R., Maduro, Q.L., Maduro, V.B., Masliello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Prasad, A., Stantripop, S., Thomas, J.W., Thomas, P.J., Touchman, J.W., Tsourounis, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L., Young, A., Zhang, L.-H. and Green, E.D.
TITLE	NISC Comparative Sequencing Initiative
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 145123)
AUTHORS	Green, E.D.
TITLE	Direct Submission
JOURNAL	Submitted (04-JUL-2001) NIH Intramural Sequencing Center, 8717
REFERENCE	Groveview Circle, Gaithersburg, MD 20877, USA
COMMENT	On Jan 30, 2002 this sequence version replaced gi:14595776.
	----- Genome Center
	Center: NIH Intramural Sequencing Center
	Center code: NISC
	Web site: <a href="http://www.nisc.nih.gov">http://www.nisc.nih.gov</a>
	Contact: <a href="mailto:nisc.mouse@nih.gov">nisc.mouse@nih.gov</a>
	----- Project Information

Center project name: apk  
Center clone name: 100N11

The sequence data in this record represents an 'enhanced' version of a Phase 2 submission. Specifically, the indicated order and orientation of each sequence contig has been established using one or more of the following: read-pair data from individual subclones, overlaps with neighboring clones, alignment with available reference sequence (e.g., human), and/or confirmation by PCR testing. In addition, the sequence assembly is based on at least 8X average coverage in Q20 bases and has been reviewed to rule out contig misassemblies, the low-quality ends of sequence with Phrap-derived quality score.

----- Summary Statistics -----

Sequencing vector: Plasmid; n/a; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 143439 bases at least Q40  
Consensus quality: 143941 bases at least Q30  
Insert size: 143000; agarose-fp  
Insert size: 144723; sum-of-contigs  
Quality coverage: 12.81x in Q20 bases; agarose-fp  
Quality coverage: 12.05x in Q20 bases; sum-of-contigs

----- NOTE: This is a 'working draft' sequence. It currently consists of 5 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.

\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved  
\* 1 5391: contig of 5391 bp in length  
\* 5392 5491: gap of unknown length  
\* 5492 7489: contig of 1998 bp in length  
\* 7490 7589: gap of unknown length  
\* 7590 89522: contig of 81933 bp in length  
\* 89523 89622: gap of unknown length  
\* 89623 92709: contig of 3087 bp in length  
\* 92710 92809: gap of unknown length  
\* 92810 145123: contig of 52314 bp in length.

#### FEATURES

Source	Location/Qualifiers
misc_feature	1..145123 /organism="Gallus gallus" /db_xref="taxon:9031" /clone="100N11"
misc_feature	1..5391 /note="assembly-fragment" vector_end:sp6 vector_side:left /note="7489" /note="assembly-fragment"
misc_feature	7590..89522 /note="assembly-fragment"
misc_feature	88183..92709 /note="assembly-fragment"
misc_feature	AC092081 clone 77B19 (center project name apl) 89623..92709 /note="assembly-fragment"
misc_feature	92810..145123 /note="assembly-fragment" vector_end:t7 vector_side:right

BASE COUNT 37387 a 33583 c 34208 g 39540 t 405 others

Query Match

8.7%; Score 62.8; DB 2; Length 145123;

Best Local Similarity 76.3%; Pred. No. 0.00044;  
Matches 90; Conservative 27; Indels 1; Gaps 1;  
QY 16 CAAGGAATGACATAGATGGCACTTAGTCCATGCTAGTTCACAGGTGA-TGGTTGG 74  
DB 71814 CAGGAATGAGAGATGGCACTGAGGACATGTTTAGTGGCTATATAGTGGTAGG 71755  
QY 75 TCAAAAGTGGACTCGATGATCTCAGAGTCTTTTCCAGCCTTATATCTATGAAT 132  
DB 71754 TGGACAGCTGGCTAGATGATCTCAGAGTCTTTTCCAGCCTTATATCTATGAAT 71697

#### RESULT 4

AC091708 124309 bp DNA linear VRT 09-AUG-2001  
Gallus gallus clone 68C5, complete sequence.  
AC091708  
AC091708.2 GI:15145070  
HTG.  
SOURCE Chicken.  
ORGANISM Gallus gallus

REFERENCE 1 (bases 1 to 124309)  
AUTHORS Aya, E.K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W.,  
Griffard, G.G., Breen, K., Brinkley, C., Brooks, S., Dietrich, N.L.,  
Graham, S., Guan, X., Gupta, J., Ho, S.-L., Idol, J.R., Karlins, E.,  
Lee, S.-Q., Legaspi, R., Lim, M., Maduro, Q.L., Maduro, V.B.,  
Maselli, C., Mastrian, S.D., McCloskey, J.C., McDowell, J.,  
Pearson, R., Pasad, A., Shevchenko, Y., Snyder, B., Stantripop, S.,  
Thomas, J.W., Thomas, P.J., Touchman, J.W., Tsurgenev, C., Vogt, J.L.,  
Walker, M.A., Wehrt, K.D., Zhang, L.-H. and Green, E.D.  
NISC Comparative Sequencing Initiative

#### REFERENCE

1 (bases 1 to 124309)

Unpublished

2 (bases 1 to 124309)

Direct Submission

Submitted (24-MAY-2001) NIH Intramural Sequencing Center, 8717

Grovemont Circle, Gaithersburg, MD 20877, USA

3 (bases 1 to 124309)

Green, E.D.

Direct Submission

Submitted (09-AUG-2001) NIH Intramural Sequencing Center, 8717

Grovemont Circle, Gaithersburg, MD 20877, USA

On Aug 9, 2001 this sequence version replaced gi:14192887.

----- Genomic Center

Center: NIH Intramural Sequencing Center

Center code: NISC

Web site: http://www.nisc.nih.gov

Contact: nisc.mouse@nhgri.nih.gov

----- Project Information

Center project name: apo

Center clone name: 068C05

-----

This sequence was finished as follows unless otherwise noted:

all regions were double-stranded, sequenced with an

alternate chemistry, or covered by high quality data

(i.e., phred quality >= 30); an attempt was made to resolve

all sequencing problems, such as compressions and repeats;

all regions were covered by at least one plasmid subclone

or more than one M13 subclone; and the assembly was confirmed

by restriction digest.

CLONE LENGTH: This sequence represents the entire insert of

this clone unless otherwise noted. If there are overlapping

clones, the overlaps are noted in the beginning and end of

the Features section.

Location/Qualifiers

1..124309

/organism="Gallus gallus"

/db\_xref="taxon:9031"

/clone="68C5"

-----

SOURCE

```
/clone_lib="GGR-G. gallus-Red Jungle Fowl
UCD001-pBelOBAC11/BI BAC library"
11962. 12098
/clone="single clone coverage"
83376. >124309
/note="this sequence is not the entire insert of clone
68C5; clone overlaps with GenBank Accession Number
AC084760 (nucleotides 1-52406) clone 65N20 (center project
name apc); this annotated segment represents overlap with
nucleotides 1-40929 of AC084760"
BASE COUNT 37444 a 25022 c 24937 g 36906 t
ORIGIN

Query Match 8.6%; Score 62; DB 5; Length 124309;
Best Local Similarity 70.3%; Pred. No. 0.00067;
Matches 83; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

QY 14 TTCAAGGAATGACTAGTGGCACTTAGTGGCCATGCTAGTTGACAAGGTGATGGTTCGAAAGTTGGACTCGA 91
Db 71279 TTCAAGGAATGACTAGTGGCACTTAGTGGCCATGCTAGTTGACAAGGTGATGGTTCGAAAGTTGGACTCGA 144
QY 74 GTCAAAAGTGGACTCGATGATCTCAGAGTCTTTTCCAGCCCTTAATATCTATGAA 131
Db 71339 GCTGATGGTTGACTAGATGCTTACAGGTCTTTTCCAGCCCTTAATATCTATGAA 71396

RESULT 5
AF076669/c 230 bp DNA linear VRT 10-JAN-2001
LOCUS
DEFINITION Zosterops lateralis chlorocephala clone 2L22, microsatellite
sequence.
ACCESSION AF076669
VERSION AF076669
KEYWORDS
PUBMED 991705
SOURCE Zosterops lateralis chlorocephala.
ORGANISM Zosterops lateralis chlorocephala.
REFERENCE 1 (bases 1 to 230)
AUTHORS Degnan,S.M., Robertson,B.C., Clegg,S.M. and Moritz,C.C.
TITLE Microsatellite primers for studies of gene flow and mating systems
in white-eyes (Zosterops)
JOURNAL Mol. Ecol. 8 (1), 159-160 (1999)
MEDLINE 99118358
PUBMED 991705
REFERENCE 2 (bases 1 to 230)
AUTHORS Degnan,S.M.
TITLE Direct Submission
JOURNAL Submitted (09-JUL-1998) zoology, University of Queensland,
Brisbane, Qld 4072, Australia
FEATURES
source
1..230
/organism="Zosterops lateralis chlorocephala"
/sub_species="chlorocephala"
/db_xref="taxon:79321"
/clone="2L22"
/note="Capricorn Silvereye"
primer_bind 14..39
repeat_region 70..87
/note="microsatellite"
/rpt_type=tandem
/rpt_unit=gt
primer_bind 71 a 51 c 42 g 66 t
BASE COUNT 71 a 51 c 42 g 66 t
ORIGIN

Query Match 8.5%; Score 61.2; DB 5; Length 230;
Best Local Similarity 71.1%; Pred. No. 0.0025;
Matches 81; Conservative 0; Mismatches 33; Indels 0; Gaps 0;
```

```
QY 32 GTGGCACTTAGTCCCATGCTAGTTGACAAGGTGATGGTTCGAAAGTTGGACTCGA 91
Db 203 GTGGCACTAGTCTGCTGCTGGGTGACAAGTTTGGGTGCTAGTGGATTCA 144
QY 92 TGATCTCAGAGTCTTTTTCAGCCCTTAATATCTATGAAATCTGTAATTTAT 145
Db 143 TGACCTTGGAGTCTTTTTCGACCTAAATGATTCATCCATGATTCATGTTAT 90

RESULT 6
AP003796 156014 bp DNA linear VRT 20-OCT-2001
LOCUS
DEFINITION Gallus gallus genomic DNA, chromosome 5, clone:26D12, complete
sequence.
ACCESSION AP003796
VERSION AP003796.2 GI:16303289
KEYWORDS HTG.
SOURCE Gallus gallus DNA, clone:26D12.
ORGANISM Gallus gallus
REFERENCE 1 (bases 1 to 156014)
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Homo sapiens genomic DNA
JOURNAL Published Only in Database (2001) In press
REFERENCE 2 (bases 1 to 156014)
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (27-JUN-2001) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC),
1-7-22 Suehiro-cho,Tsrumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:hattori@gs.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT On Oct 19, 2001 this sequence version replaced gi:15208285.
This work was done in collaboration with Yokomine, T. and Sakaki,
H.
National Institute of Genetics
Human Genetics Division
1-111, Yata, Mishima, Shizuoka, JAPAN
zip: 411-8540
e-mail: hisasakielab.nig.ac.jp.
LOCATION/Qualifiers
1..156014
/organism="Gallus gallus"
/db_xref="taxon:9031"
/chromosome="5"
/map="5"
/clone="26D12"
BASE COUNT 40933 a 37714 c 36721 g 40598 t 48 others
ORIGIN

Query Match 8.2%; Score 59.4; DB 5; Length 156014;
Best Local Similarity 64.6%; Pred. No. 0.0024;
Matches 104; Conservative 0; Mismatches 56; Indels 1; Gaps 1;

QY 1 CCGGTCGGAGGTTTCAAGGAATGACTAGTGGGCACCTAGTCCATGGCTGATTGAC 60
Db 39597 CCGCTTCCATGGGCAAGCAAGGATAGATGGTATTGAAGGACATGG-CTAGAGGGC 39655
QY 61 AAGTGATGTTGGTCAAAAGTTGGACTCGATGATCTCAGAGTCTTTTCCAGCCCTTAAT 120
Db 39656 ACGTGGGATGGCTGACAGTTGGACTAGATGATCTTACTGGTCCATCCACCCCTATT 39715
QY 121 AATTCTATGAATCTCTAATTTTATCTTCATGATCTTTTGAG 161
Db 39716 GTTCTATTGATCTCTGATTTCTTAAGCTGAGTGTGCGAGGAG 39756

RESULT 7
```



BASE COUNT	8818 a	6981 c	7243 g	8750 t	1 others
ORIGIN					
Query Match		8.0%	Score 57.8;	DB 5;	Length 31793;
Best Local Similarity		68.6%;	Pred. No. 0.0066;		
Matches	94;	Conservative	0;	Mismatches	42; Indels 1; Gaps 1;
Qy	4	GTCGCGAGGTTTCAAGCAATGACTAGATGTGGCACTTAGTGCATGTGGCTAGTGGTCTAGTTGACAAAG	63		
Db	9714	GATGTGGGTGTTCAAGAAATGCTAGACATGCTATTTAGGGACATGGCTTAGTGGCAAT	9773		
Qy	64	GTGA-TGGTGTGTTCAAAAGTTGGACTCGATGATCTCAGAGTCTTTTTTCCAGCCTTAATAA	122		
Db	9774	ATTAGTGTGAGTGTAGTACTAAATGCTGTGGGGTCTTTTCCAGCCTTAATGA	9833		
Qy	123	TTCTATGAATTCCTGTA	139		
Db	9834	TTCTATGACTAATAGAA	9850		
RESULT	8				
AP003795/c					
LOCUS		AP003795	187259 bp	DNA	linear
DEFINITION					Gallus gallus genomic DNA, chromosome 5, clone:192C9, complete
ACCESSION		AP003795			
VERSION		AP003795.2	GI:16303288		
KEYWORDS		HTG			
SOURCE		Gallus gallus			
ORGANISM		Gallus gallus DNA, clone:192C9.			
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
COMMENT					
FEATURES					
source					
BASE COUNT	53136 a	43577 c	41220 g	49326 t	
ORIGIN					
Query Match		7.9%	Score 57;	DB 5;	Length 187259;
Best Local Similarity		65.1%;	Pred. No. 0.0076;		
Matches	84;	Conservative	0;	Mismatches	45; Indels 0; Gaps 0;
Qy	8	GGAGGTTTCAAGCAATGACTAGATGTGGCACTTAGTGCATGTGGCTAGTGGTCTAGTTGACAAAGTGA	67		

```
Db 64635 GGAGGTGCTCAGGAACATGAGATATAGCACTGAGGACGTGGTCAGTGAGCAATATTGG 64576
QY 68 TGGTTGGTCAAAAGTTGGACTCGATGATCTCAGAGTCTTTTCCAGCCCTTAATAATTCTA 127
Db 64575 CGATAGTAGACTGTGGACCTGATGATCTTACAGATGCTTCCACCTTAATGATTITA 64516
QY 128 TGAATTCCTG 136
Db 64515 TGATTGCTG 64507

RESULT 9
AC084760
LOCUS      109569 bp      DNA      linear      VRT 26-JAN-2001
DEFINITION      Gallus gallus clone 65N20, complete sequence.
ACCESSION      AC084760
VERSION      AC084760.2  GI:12545295
KEYWORDS      HTG.
SOURCE      chicken.
ORGANISM      Gallus gallus
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
              Phasianinae; Gallus
              1 (bases 1 to 109569)
              Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W.,
              Bouffard, G.G., Brinkley, C., Brooks, S., Dietrich, N.L., Granite, S.,
              Guan, X., Gupta, J., Ho, S.-L., Idol, J.R., Karlins, E., Lee-Lin, S.-Q.,
              Legaspi, R., Lim, M., Maduro, O.L., Maduro, V.B., Masiello, C.,
              Mastrian, S.D., McCloskey, J.C., McDowell, J.J., Pearson, R., Prasad, A.,
              Shevchenko, Y., Snyder, B., Stantropow, S., Thomas, J.W., Thomas, P.J.,
              Tongson, E.E., Touchman, J.W., Tsurgeon, C., Vogt, J.L., Walker, M.A.,
              Wetherby, K.D., Zhang, L.-H. and Green, E.D.
              NISC Comparative Sequencing Initiative
              Unpublished
              2 (bases 1 to 109569)
              Green, E.D.
              Direct Submission
              Submitted (15-NOV-2000) NIH Intramural Sequencing Center, 8717
              Grovemont Circle, Gaithersburg, MD 20877, USA
              3 (bases 1 to 109569)
              Green, E.D.
              Direct Submission
              Submitted (26-JAN-2001) NIH Intramural Sequencing Center, 8717
              Grovemont Circle, Gaithersburg, MD 20877, USA
              On Jan 26, 2001 this sequence version replaced gi:11178075.
              ----- Genome Center
              Center: NIH Intramural Sequencing Center
              Center code: NISC
              Web site: http://www.nisc.nih.gov
              Contact: nisc_mouse@hgri.nih.gov
              ----- Project Information
              Center project name: apc
              Center clone name: 065N20

This sequence was finished as follows unless otherwise noted:
all regions were double-stranded, sequenced with an
alternate chemistry, or covered by high quality data
(i.e., phred quality >= 30); an attempt was made to resolve
all sequencing problems, such as compressions and repeats;
all regions were covered by at least one plasmid subclone
or more than one M13 subclone; and the assembly was confirmed
by restriction digest.

CLONE LENGTH: This sequence represents the entire insert of
this clone unless otherwise noted. If there are overlapping
clones, the overlaps are noted in the beginning and end of
the Features section.
              Location/Qualifiers
              1..109569
              /organism="Gallus gallus"
              /db_xref="taxon:9031"
              /clone="65N20"
              /clone_lib="GGR-G. domesticus-PECBACI/HIII BAC library"

FEATURES
source
```

```
misc_feature      720..770
                  /note="single clone coverage"
misc_feature      56700..56773
                  /note="single clone coverage"
BASE COUNT      32522 a 22972 c 23082 g 30993 t
ORIGIN

Query Match      7.7%; Score 55.4; DB 5; Length 109569;
Best Local Similarity 64.3%; Pred. No. 0.018;
Matches 83; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

QY 2 CCGTCGCGAGGTTTCAAGAACTAGTAGTGCGGCACTTAGTGCCATGCTAGTTGACA 61
Db 81900 CCCCTGGGGTCTCGAGAAACAAGCAGATGTGGCACAGGAGCGTGTTCAGTACGCA 81959
QY 62 AGGTGATGTTGGTCAAAAGTTGGACTCGATGATCTCAGAGTCTTTTCCACGCTTAATA 121
Db 81960 CTGTGGGAAGGGTTGATGTTGGACTAGTGATCTTAGTGCTCTTTCCAACTTGCTG 82019
QY 122 ATTCTATGA 130
Db 82020 ATTCATCA 82028

RESULT 10
AP003796/c
LOCUS      156014 bp      DNA      linear      VRT 20-OCT-2001
DEFINITION      Gallus gallus genomic DNA, chromosome 5, clone:26D12, complete
                  sequence.
ACCESSION      AP003796
VERSION      AP003796.2  GI:16303289
KEYWORDS      HTG.
SOURCE      Gallus gallus DNA, clone:26D12.
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
              Phasianinae; Gallus
              1 (bases 1 to 156014)
              Fujiyama, A., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.
              Hattori, M., Ishii, K., Toyoda, A., Taylor, T.D., Hong-Seog, P.,
              Homo sapiens genomic DNA
              Published Only in Database (2001) In press
              2 (bases 1 to 156014)
              Hattori, M., Ishii, K., Toyoda, A., Taylor, T.D., Hong-Seog, P.,
              Fujiyama, A., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.
              Direct Submission
              Submitted (27-JUN-2001) Masahira Hattori, The Institute of Physical
              and Chemical Research (RIKEN), Genomic Sciences Center (GSC); Japan
              1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
              (E-mail: hattori@gsc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/,
              Tel: 81-45-503-9111, Fax: 81-45-503-9170)
              On Oct 19, 2001 this sequence version replaced gi:15208285.
              This work was done in collaboration with Yokomine, T. and Sasaki,
              H.
              National Institute of Genetics
              Human Genetics Division
              1-111, Yata, Mishima, Shizuoka, JAPAN
              zip: 411-8540
              e-mail: hisasaki@lab.nig.ac.jp.
              Location/Qualifiers
              1..156014
              /organism="Gallus gallus"
              /db_xref="taxon:9031"
              /chromosome="5"
              /map="5"
              /clone="26D12"
              /clone="26D12"

BASE COUNT      40933 a 37714 c 36721 g 40598 t 48 others
ORIGIN

Query Match      7.6%; Score 55.2; DB 5; Length 156014;
Best Local Similarity 71.7%; Pred. No. 0.019;
```

[illegible]

Query Match	7.6%;	Score 55;	DB 5;	Length 5805;
Best Local Similarity	66.4%;	Pred. NO.	0.034;	
Matches	79;	Conservative	0;	Mismatches 40; Indels 0; Gaps 0;
QY	14	TTCACGAATGACTAGATGGCGCTAGTGCCTAGTGCTAGTTGGACAAGGTGATGGTTG	73	
Db	299	TTCACGAAAGCGTGGATGTAGTAATTAGCACACAGTTCATTGGGAAATPACTGCCGGTAG	358	
QY	74	GTCAAAAGTGGCACTCGATCTCAGAGTTTTTTCCAGCCCTTAATAATCTATGTAAT	132	
Db	359	GTGGATGGTGGACATGACAGTCTTAGAGATCTTTTCCAACCTTAATGCTTCTATGATT	417	
RESULT	12			
AX345619				
LOCUS	AX345619	12007 bp	DNA	linear
DEFINITION	Sequence 690 from Patent WO0200928.			
ACCESSION	AX345619			
VERSION	AX345619.1	GI:18493505		
KEYWORDS	synthetic construct.			
SOURCE	synthetic construct			
ORGANISM	artificial sequence.			

[illegible]

RESULT	13
AC091708/c	
LOCUS	linear VRT 09-AUG-2001
DEFINITION	Gallus gallus clone 68C5, complete sequence.
AC091708	124309 bp DNA
VERSION	AC091708.2 GI:15145070
KEYWORDS	HTG.
SOURCE	chicken.
ORGANISM	Gallus gallus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
REFERENCE	Phasianinae; Gallus. 1 (bases 1 to 124309)
AUTHORS	Avelle,K., Beckstrom-Sternberg,S.M., Benjamin,B., Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S., Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Ho,S.-L., Idol,J.R., Karlins,E., Lee-Lin,S.-Q., Legaspi,R.K., Lim,H.S., Maduro,V.B., Mastello,C., Mastrian,S.D., McCloskey,J.C., McDowell,J., Pearson,R., Prasad,A., Shevchenko,Y., Snyder,B., Stantropop,S., Thomas,J.W., Thomas,P.J., Touchman,J.W., Tsurgeon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Zhang,L.H. and Green,E.D.
TITLE	NISC Comparative Sequencing Initiative
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 124309)
AUTHORS	Green,E.D.
TITLE	Direct Submission

JOURNAL Submitted (24-MAY-2001) NIH Intramural Sequencing Center, 8717  
 GroveMont Circle, Gaithersburg, MD 20877, USA  
 REFERENCE 3 (bases 1 to 124309)  
 AUTHORS Green, E.D.  
 TITLE Direct Submission  
 JOURNAL Submitted (09-AUG-2001) NIH Intramural Sequencing Center, 8717  
 GroveMont Circle, Gaithersburg, MD 20877, USA  
 COMMENT On Aug 9, 2001 this sequence version replaced gi:14192887.  
 ----- Genome Center  
 Center: NIH Intramural Sequencing Center  
 Center code: NISC  
 Web site: <http://www.nisc.nih.gov>  
 Contact: [nisc\\_mouse@nhgri.nih.gov](mailto:nisc_mouse@nhgri.nih.gov)  
 ----- Project Information  
 Center project name: apo  
 Center clone name: 068C05

This sequence was finished as follows unless otherwise noted:  
 all regions were double-stranded, sequenced with an  
 alternate chemistry, or covered by high quality data  
 (i.e., phred quality >= 30); an attempt was made to resolve  
 all sequencing problems, such as compressions and repeats;  
 all regions were covered by at least one plasmid subclone  
 or more than one M13 subclone; and the assembly was confirmed  
 by restriction digest.

CLONE LENGTH: This sequence represents the entire insert of  
 this clone unless otherwise noted. If there are overlapping  
 clones, the overlaps are noted in the beginning and end of  
 the Features section.

FEATURES  
 source  
 1..124309  
 /organism="Gallus gallus"  
 /db\_xref="taxon:9031"  
 /clone="68C5"  
 /clone\_lib="GGR-G. gallus-Red Jungle Fowl"  
 /cd001-pbeloBAC11/BI BAC library"  
 11962..12098  
 /note="single clone coverage"  
 83376..>124309  
 /note="this sequence is not the entire insert of clone  
 68C5; clone overlaps with GenBank Accession Number  
 AC084760 (nucleotides 1-52406) clone 65N20 (center project  
 name apc); this annotated segment represents overlap with  
 nucleotides 1-40929 of AC084760"  
 BASE COUNT 37444 a 25022 c 24937 g 36906 t  
 ORIGIN

Query Match 7.6%; Score 54.6; DB 5; Length 124309;  
 Best Local Similarity 64.8%; Pred. NO. 0.026;  
 Matches 81; Conservative 0; Mismatches 44; Indels 0; Gaps 0;  
 QY 8 GGAGGTTTCAAGGATGACTAGATGGGCACCTAGTGGCCATGGTCTAGTGTGACAAGGTGA 67  
 Db 11538 GGATGATTATCAACAACTGAGGAGGCGGACGAACATGGTTATTGGACGTGATGG 11479  
 QY 68 TGGTGGTCAAAAGTTGGACTCGATGATCTCAGAGTTTTTTTCCAGCCCTAATAATCTTA 127  
 Db 11478 TGATGGCTGATGATGAACTGGGTGATCTTAGTGGTCTTTCCAACTTAATGATCGA 11419  
 QY 128 TGAAT 132  
 Db 11418 TGATT 11414

RESULT 14  
 AF111666/c  
 LOCUS AF111666 314 bp DNA linear VRT 19-SEP-1999  
 DEFINITION Meleagris gallopavo clone TUCA930 microsatellite sequence.  
 ACCESSION AF111666  
 VERSION AF111666.1 GI:5911090  
 KEYWORDS

SOURCE turkey.  
 ORGANISM Meleagris gallopavo  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Archosauria; Aves; Neognathae; Galliformes; Meleagridae;  
 Meleagris.  
 REFERENCE 1. (bases 1 to 314)  
 AUTHORS Smith, E.J., Nahason, S., Shi, L., Drummond, P., Zahorchak, R. and  
 Foster, J.C.  
 TITLE Genomic DNA sequence from a DNA library from turkey microsatellites  
 JOURNAL Unpublished  
 REFERENCE 2. (bases 1 to 314)  
 AUTHORS Smith, E.J., Nahason, S., Shi, L., Drummond, P., Zahorchak, R. and  
 Foster, J.C.  
 TITLE Direct Submission  
 JOURNAL Submitted (08-DEC-1998) Comparative Genomics, Tuskegee University,  
 109 Milbank Hall, Tuskegee, AL 36088, USA  
 FEATURES  
 source  
 1..314  
 /organism="Meleagris gallopavo"  
 /db\_xref="taxon:9103"  
 /clone="TUCA930"  
 repeat\_region 1..314  
 /note="microsatellite"  
 /rpt\_type=tandem  
 BASE COUNT 105 a 84 c 56 g 69 t  
 ORIGIN  
 Query Match 7.5%; Score 54.2; DB 5; Length 314;  
 Best Local Similarity 71.4%; Pred. NO. 0.077; 33; Indels 1; Gaps 1;  
 Matches 85; Conservative 0; Mismatches 33; Indels 1; Gaps 1;  
 QY 14 TTCAAGGAATGACTAGATGGCCTAGTGGCCATGGTCTAGTGTGACAAGGTGATGGTGG 73  
 Db 214 TTCAAGAACTATGTGGATGTTACACTGAGACATGGT-TAGTGGCAGTGGTATGATGG 156  
 QY 74 GTCAAAAGTTGGACTCGATGATCTCAGAGTTTTTTTCCAGCCCTTAATAATCTTCAAT 132  
 Db 155 ATTGATGGTTGAATTAGATGATCTTAGTGGTCTTTTCCAACTTAATGTTCTATGATT 97  
 RESULT 15  
 AC092081/c 165108 bp DNA linear HTG 30-JAN-2002  
 LOCUS Gallus gallus clone 77D19, WORKING DRAFT SEQUENCE, 5 ordered  
 DEFINITION pieces.  
 ACCESSION AC092081  
 VERSION AC092081.2 GI:18425283  
 KEYWORDS HTG; HTGS\_PHASE2; HTGS\_DRAFT.  
 SOURCE chicken.  
 ORGANISM Gallus gallus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
 Phasianinae; Gallus.  
 REFERENCE 1 (bases 1 to 165108)  
 AUTHORS Akhter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B.,  
 Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,  
 Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P.,  
 Ho, S.-L., Idol, J.R., Karlins, E., Laric, P., Lee-Lin, S.-O.,  
 Legaspi, R., Maduro, O.L., Maduro, V.B., Masiello, C., Maskeri, B.,  
 Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Prasad, A.,  
 Stantripop, S., Thomas, J.W., Thomas, P.J., Touchman, J.W.,  
 Tsurgon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,  
 Young, A., Zhang, L.-H. and Green, E.D.  
 TITLE NISC Comparative Sequencing Initiative  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 165108)  
 AUTHORS Green, E.D.  
 TITLE Direct Submission  
 JOURNAL Submitted (20-JUN-2001) NIH Intramural Sequencing Center, 8717  
 GroveMont Circle, Gaithersburg, MD 20877, USA  
 COMMENT On Jan 30, 2002 this sequence version replaced gi:14495366.  
 ----- Genome Center





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 2, 2002, 13:28:39 ; Search time 9940.62 Seconds  
(without alignments)  
1091.457 Million cell updates/sec

Title: US-08-973-363-1

Perfect score: 723

Sequence: 1 CCGGTCGGAGGTTTCAAGS.....CTTGAACCTCCGACCGG 723

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_estl:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_gss:\*  
13: em\_gss\_hum:\*  
14: em\_gss\_inv:\*  
15: em\_gss\_pln:\*  
16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	78.6	10.9	368	9	AL585441
2	63.2	8.7	974	12	CNS001IT
3	61.2	8.5	630	9	AJ398678
4	59.2	8.2	1101	12	CNS00262
5	58	8.0	804	12	AG077527
6	57.2	7.9	1112	12	GGA340114
7	57	7.9	883	9	AL575329
8	56.8	7.9	585	9	AL588169
9	56.8	7.9	619	9	AL514935
10	56.2	7.8	716	9	AL584261
11	55.8	7.7	799	9	AJ393303
12	55.8	7.7	1147	12	B13042
13	55.4	7.7	919	12	CNS005RL
14	55	7.6	411	12	GGA246046
15	55	7.6	1101	12	CNS0100X
16	54.8	7.6	1101	12	CNS00GDR
17	54.6	7.6	1101	12	CNS0039G

c 18	54.4	7.5	589	12	CNS0301Z
c 19	54.4	7.5	803	12	BL1974
c 20	54	7.5	581	12	CNS01UP2
c 21	53.8	7.4	713	12	AO576566
c 22	53.8	7.4	821	12	CNS0387F
c 23	53.6	7.4	1050	10	BM415695
c 24	53.2	7.4	576	12	CNS035N7
c 25	53.2	7.4	1101	12	CNS00BGD
c 26	53	7.3	756	3	BL644558
c 27	52.6	7.3	1101	12	CNS00L72
c 28	51.8	7.2	605	10	BM491757
c 29	51.8	7.2	953	12	AG077643
c 30	51.8	7.2	987	12	CNS004EE
c 31	51.8	7.2	1185	10	BF273407
c 32	51.6	7.1	911	12	CNS00D2S
c 33	51.6	7.1	1015	12	CNS00IUO
c 34	51.2	7.1	717	12	CNS06ZNC
c 35	51	7.1	959	12	CNS00655
c 36	51	7.1	1025	12	CNS014J2
c 37	50.8	7.0	915	10	BG520365
c 38	50.8	7.0	1200	12	CNS016CO
c 39	50.6	7.0	948	10	BM415494
c 40	50.4	7.0	615	9	AJ981623
c 41	50.4	7.0	854	12	CNS012CM
c 42	50.4	7.0	909	12	CNS02083
c 43	50.4	7.0	1146	12	CNS02162
c 44	50.2	6.9	620	10	BL389975
c 45	50.2	6.9	1101	12	CNS01807

#### ALIGNMENTS

RESULT 1

AL585441

LOCUS

DEFINITION

ACCESION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

AL585441 368 bp mRNA linear EST 28-FEB-2001  
AL585441 Stratagene Chick Embryo Lambda cDNA Library (\* 937405)  
Gallus gallus cDNA clone ROS023H05, mRNA sequence.

AL585441

AL585441

AL585441.1 GI:13164174

EST.

chicken.

Gallus gallus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Archosauria; Aves; Neognathae; Galliformes; Phasianidae;

Phasianinae; Gallus.

1 (bases 1 to 368)

Murray, F. Chick Embryo Lambda cDNA Library

Stratagene (2001)

Unpublished (2001)

Contact: Frazer Murray

Dept. Genomics and Bioinformatics

Roslin Institute

Roslin, Midlothian, EH25 9PS, UK

Tel: +44 (0)131 527 4200

Fax: +44 (0)131 440 0434

Email: frazer.murray@bbsrc.ac.uk

Seq primer: T3.

Location/Qualifiers

1. 368

/organism="Gallus gallus"

/db\_xref="taxon:9031"

/clone="ROS023H05"

/clone.lib="Stratagene Chick Embryo Lambda cDNA Library (\* 937405)"

/tissue.type="Embryo"

/dev.stage="5 days old"

/lab\_host="SOLR cells (kanamycin resistant)"

/note="Vector: pBLUESCRIPT SK; Site: 1: EcoRI; Site: 2: XhoI

; Cloned unidirectionally. Primer: Oligo dt. uni-ZAP XR

vector. Average insert size: 1.5kb.; 5' adaptor sequence:

5' GAATTCCGACGAG 3'; 3' adaptor sequence: 5'

CTCGAGTTTTTTTTTTTTTTT 3"









RESULT 8

AL588169  
ALOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
EST.  
SOURCE  
ORGANISM

AL588169 585 bp mRNA linear EST 02-MAR-2001  
AL588169 BP Chicken Brain Library Gallus gallus cdNA clone  
RS068E07, mRNA sequence.  
AL588169  
AL588169.1 GI:13193203  
EST.  
SOURCE  
ORGANISM

Gallus gallus  
chicken.  
Gallus gallus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
Phasianinae; Gallus.  
1 (bases 1 to 585)  
Murray, F.  
BP Chicken Brain Library  
Unpublished (2001)  
Contact: Frazer Murray  
Dept. Genomics and Bioinformatics  
Roslin Institute  
Roslin, Midlothian, EH25 9PS, UK  
Tel: +44 (0)131 527 4200  
Fax: +44 (0)131 440 0434  
Email: frazer.murray@bbsrc.ac.uk  
GCGCGCGCTTTTTTTTTTTTTTTT 3' Poly A RNA purchased from Clontech  
(\*6854-  
Seq primer: T7.  
Location/Qualifiers  
1..585

FEATURES  
source

```

/align="global", alias="aligns"
/db_xref=taxon:9031"
/clone="ROS08E07"
/clone_lib="Bp Chicken Brain Library"
/tissue_type="brain"
/dev_stage="Unknown"
/lab_host="DH10B"
/note="vector: pSPORT1; Site_1: NotI; Site_2: SalI; Cloned
unidirectionally. Primer: Oligo dt. 5' adaptor sequence:
5' TCGACCTCGAG 3' ; 3' adaptor sequence: 5'
GCGCGCGGTTTTTTTTTTTTTTTTT 3'; Poly A RNA purchased from
Clontech (*6854-1)"
177 a 103 c 123 g 182 t
BASE COUNT
ORIGIN
Query Match 7.9%; Score 56.8; DB 9; Length 585;
Best Local Similarity 72.1%; Pred. No.: 17;
Matches 88; Conservative 0; Mismatches 32; Indels 2; Gaps 1;
Qy 14 TTCAGGAAGTACATAGCTGGCAGTTAGTGCCTAGTGTGACAGAGTGATGGTTG 73
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Ddb 315 GTGGAAGAAATCTCAGATGTGTCTCTGAGGGATATGCTTTAGTGGGCAA --TGTGTATAG 372
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 74 GTCAAAAGTTGGACTCGATGATCTCAGAGTTTTTTTCCAGCCTTAATAATTCTATGAATT 133
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Ddb 373 GTGGATGGCTGGACTGGATGATCTTAGAGTCTTTTCCAACTTGGTGATTCTATGATCT 432
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 134 CT 135
||
Db 433 CT 434

```

RESULT	9
LOCUS	AL514935
DEFINITION	AL514935 LTR_NFL006_P12 Homo sapiens cDNA clone CL0B012ZD08 3 prime, mRNA sequence.
ACCESSION	AL514935
VERSION	AL514935.1
KEYWORDS	GI:12778428
SOURCE	EST.
ORGANISM	human.
	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
1 (bases 1 to 619)  
Li.W.B., Gruber,C., Jessee,J. and Polayes,D.  
Full-length cDNA libraries and normalization  
Unpublished (2001)  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.  
Location/Qualifiers  
1. 619  
/organism="Homo sapiens"  
/dbx\_ref="taxon:9606"  
/clone="ClDB80122008"  
/clone\_lib="LTI\_NFL006\_PL2"  
/tissue\_type="placenta"  
/note=Vector: pCMVSPORT 6; Site\_1: NotI; 1st strand cDNA  
was primed with a NotI-oligo(dT) primer. Five prime end  
enriched, double-stranded cDNA was digested with Not I and  
cloned into the Not I and Eco RV sites of the pCMVSPORT 6  
vector. Library was normalized. Library was constructed by  
Life Technologies. Contact : Feng Liang Life Technologies,  
a division of Invitrogen 9800 Medical Center Drive  
Rockville, Maryland 20850, USA Fax : (1) 301 610 8371  
Email : fliang@lifestech.com URL :  
http://fulllength.invitrogen.com  
52 a 8 c 10 g 514 t 35 others  
BASE COUNT  
ORIGINAL

Query Match	7.9%;	Score 56.8;	DB 9;	Length 619;
Best Local Similarity	43.8%;	2; Mismatches	226;	Indels 0; Gaps 0;
Matches 178;	Conservative			
103	TTTTTTTCCAGCCTTAATAATCTATGAATCTGTAAATTTATTTATTCGTATGATCTTTTGGAC	162		
63	TT	122		
163	GAAGTTTCTTTGGGGATTTTAGTTTGGTTTCCCTGTCACTGTCTTCTTCTTCCTCGAAAC	222		
123	TT	182		
223	ACTTTCATTTCCACATGAGAAATGCTGTATTTGTCAGGTTACAGTAGTCAGTACGCTG	282		
183	TT	242		
283	CTTAGAAGTAGTGAGAAACATTTAGGGAATACTGGAGTGAAGCAACACAGCTGGTACTG	342		
243	NNGGGGGNANTNNNNNTTTTTTTTTTTTTTTTTNNNNNTNNKKAATAAAAAAAAAAAAAA	302		
343	CCAAACTGTACGTTGGGATTTTGAGGACGACAGATGTATATAAATTTGCTTTAATGAT	402		
303	AAAAAANAATTT	362		
403	ATCTGCGCCCTGCGCTTCCATTAATGCTGTGTTTATGAACCACTCTTTTTTTTTTTTTTT	462		
363	TT	422		
463	TTTTTTTTTGGCTTCTTCATATCCCTGGTAATGAGTTAATGCATTT	508		
423	TT	468		

RESULT	10
AL584261	
DEFINITION	716 bp mRNA linear
AL584261	Stratagene Chick Embryo Lambda cDNA library (* 937405)
Gallusgallus	cDNA clone RGS006G07, mRNA sequence..
AL584261	
ACCESSION	GI:13162992
VERSION	EST
KEYWORDS	Chicken.
SOURCE	Gallus gallus
ORGANISM	Chickens

Martinistr. 52, 20251 Hamburg, Germany  
Email: <http://genetics.hpi.uni-hamburg.de/dt40est.html>  
Location/Qualifiers  
1..799  
/organism="Gallus gallus"  
/strain="CB"  
/db\_xref="taxon:9031"  
/clone="15a5r1"  
/clone\_lib="dkfz426"  
/tissue\_type="Bursa of Fabricius"  
238 a 145 c 135 q 274 t 7 others

Query Match	7.7%	Score 55.8	DB 9	Length 799	
Best Local Similarity	73.4%	Pred. No. 20			
Matches 69	Conservative	0	Mismatches 25	Indels 0	Gaps 0
QY	36	CACHTAGTGCACATGGCTTAGTGACAGAGTGATGGTTGGTCAAAAGTTGGACTCGATGAT	95		
Db	144	CAAGATTTTACATGGTATATGGACATGGTGGTGATAGGTCAAAGTTGGACTAGGTGAT	85		
QY	96	CTCAGAGTTTTTTTTCCAGCCCTTAATAATCTATG	129		
Db	84	CTTAGTACTTTTATCCNNCCNTAATGATCTATG	51		

BI3042	BI3042	T30M24-Sp6.1	1147 bp	DNA	linear	GSS 14-MAY-1997			
LOCUS	DEFINITION	Arabidopsis thaliana genomic clone T30M24, DNA sequence.							
ACCESSION	VERSION	BI3042.1	GI:2094174						
KEYWORDS	GSS.								
SOURCE	thale cress.								
ORGANISM	Arabidopsis thaliana								
	eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids. 1 (bases 1 to 1147)								
REFERENCE	Feng, J., Dewar, K., Buehler, E., Kim, C., Li, Y., Shinn, P., Sun, H. and Ecker, J.								
AUTHORS									

TITLE BAC End Sequences at ATGC  
 JOURNAL Unpublished (1997)  
 COMMENT Other\_GSSs: T30M24-Sp6, T30M24-Sp6.2  
 Contact: Ecker J.  
 Arabidopsis Thaliana Genome Center  
 University of Pennsylvania  
 Dept. of Biology, University of Pennsylvania, Philadelphia, PA  
 19104  
 Tel: 215-898-9384  
 Fax: 215-898-8780  
 Email: jecker@atgenome.bio.upenn.edu  
 Seq primer: Sp6  
 Class: BAC ends  
 High quality sequence start: 289  
 High quality sequence stop: 292.

FEATURES	source
Location/Qualifiers	1. .1147
	/organism="Arabidopsis thaliana"
	/strain="Columbia"
	/db_xref="taxon:3702"
	/clone="n30M24"
	/clone_lib="TAMU"
	/sex="hermaphrodite"
	/note="Vector: BelOBACII; Site_1: HindIII; Site_2: HindIII"
	; Produced by Rod Wing"
BASE COUNT	3 a 4 c 46 g 553 t 541 others
ORIGIN	
Query Match	7.7%; Score 55.8; DB 12; Length 1147;

Best Local Similarity 26.2%; Pred. No. 16;  
Matches 120; Conservative 0; Mismatches 338; Indels 0; Gaps 0;

QY 51 TCAGTGTGACAGGTGATGGTGGTCAAAAGTGGACTGATGATCTCAGAGTTTTC 110  
DB 286 TTTTNTTNNNT 345  
QY 111 CAGCCTTAATAATCTGTAATTCGTAATTTATCTGATCTTTTTCAGCGAAGTTG 170  
DB 346 NNTNNNT 405  
QY 171 TTTGGGATTTAGTTGGTTCCTCCGTCACATGTTTCTTTCCTGGAACGATTTTCAT 230  
DB 406 TTTTNTTNNNT 465  
QY 231 TTGCAACATGAGATTCCTGATTTGTCAGGTTACAAGTAGTCAATGGCTGCTTAGAAG 290  
DB 466 TNN 525  
QY 291 TAGTGAGAAACATTTAGGAAATCTGCGAGTGAAGCAACACAGTGGTACTGCCAACTG 350  
DB 526 NNN 585  
QY 351 TAGCTTTGGGATTTGAGGAGCCACAGAGTTGTATATAAATTTCTTTAATGATATCTGCC 410  
DB 586 NNN 645  
QY 411 CTTGCTCTCCATTAATGCTGTTTATGAAACACATCTTTTNTTNTTNTTNTTNTTNN 470  
DB 646 NNTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTT 705  
QY 471 GCGTCTTCATATCTGCTGTAATGAGTAAATGATTT 508  
DB 706 NTNTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTT 743

RESULT 13  
CNS005RL/c 919 bp DNA linear GSS 03-JUN-1999  
LOCUS Drosophila melanogaster genome survey sequence T7 end of BAC #  
DEFINITION BACRI2F23 of RPI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.  
ACCESSION AL061409  
VERSION AL061409.1 GI:4943512  
KEYWORDS GSS.  
SOURCE fruit fly.  
ORGANISM Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
REFERENCE 1 (bases 1 to 919)  
AUTHORS Genoscope.  
TITLE Direct Submission  
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : Bp 191 91006 Evry cedex - FRANCE (E-mail : segref@genoscope.cns.fr - Web : www.genoscope.cns.fr)  
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mamoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's p1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).  
FEATURES  
source 1. .919

/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/clone\_lib="RPI-98"  
/clone="BACRI2F23"  
/note="end : 17"  
BASE COUNT 279 a 124 c 93 g 197 t 226 others  
ORIGIN

Query Match 7.7%; Score 55.4; DB 12; Length 919;  
Best Local Similarity 32.4%; Pred. No. 21;  
Matches 110; Conservative 77; Mismatches 151; Indels 2; Gaps 1;

QY 363 TTGAGGACCCACAGAGTTGATATAAATTTTATGATATCTGCCCTGCCAT 422  
DB 775 KKGKWTWTAAATWDTKAWAARDAATRTWDTAAGKKKKTKKWAATWGATKAWA 716  
QY 423 TAATGCTGTGTTTATGAAACCACTCTTTTNTTNTTNTTNTTNTTNTTNTTNTT 482  
DB 715 TWTKKKKKDKTKWKWTWDTNAWWTWTNTTNTTNTTNTTNTTNTTNTTNTTNTT 656  
QY 483 TCCTGTGTTGTAAGTAAATGCAATTTTAGAAGCACATGGCAGAACTAGGAGATCTGTGGAT 542  
DB 655 TTTTNTTAT - ATGTDKAAGTAMWTTTAAATTTNGTNGKGGKKKKGGKKKG 598  
QY 543 GACAGTGTGACAGGAGCTCTGAATTTTATAGATAAATCTAGAGTGGAAACAGAAATCT 602  
DB 597 KKGKTKKKKKKKKKKDDTDDDDWWTTATGATGATGAAAGGAAATGAAAGGWT 538  
QY 603 GAGGCTAGTCTTCTGACCTGACCTGAAATTTTCTGAGATATTTCAAGACTACATAGT 662  
DB 537 AWTGTTGAWTTTCTGGNAGAAATTTTCTGTTATATATATATATNTTNTTCTCT 478  
QY 663 TGTGCTTTGAGGAAAATAAATGTTTAAGTGTCTCCAT 702  
DB 477 TTNGTWTTAACCCCATATWTTAATNAGNATGTCGCAAT 438

RESULT 14  
GGA246046/c 411 bp DNA linear GSS 21-JUN-1999  
LOCUS Gallus gallus anonymous sequence from cosmid mapping to chromosome 7 (Cosmid 3 - Contig 12), genomic survey sequence.  
DEFINITION GGA246046  
ACCESSION AJ246046.1 GI:5139373  
VERSION AJ246046  
KEYWORDS GSS; genome survey sequence.  
SOURCE chicken.  
ORGANISM Gallus gallus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.  
REFERENCE 1 (bases 1 to 411)  
AUTHORS Smith, J., Bruley, C.K., Paton, I.R., Law, A.S., Masabanda, J., Waddington, D., Fries, R. and Burt, D.W.  
TITLE Integration of the genetic and physical maps of the chicken macrochromosomes  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 411)  
AUTHORS Law, A.S.  
JOURNAL Direct Submission  
TITLE Submitted (14-JUN-1999) Law A.S., Division of Molecular Biology, Roslin Institute, Roslin, Midlothian EH25 9PS, Scotland, UK  
FEATURES  
source 1. .411  
/organism="Gallus gallus"  
/db\_xref="taxon:9031"  
/chromosome="7"  
/map="7q"  
BASE COUNT 105 a 127 c 71 g 108 t  
ORIGIN

Query Match 7.6%; Score 55; DB 12; Length 411;

```
Best Local Similarity 66.4%; Pred. No. 38;
Matches 79; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 14 TTCAAGGAAGTACTAGATGGCATTAGTGGCATGCTCTAGTTGACAAGGTGATGGTTG 73
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 187 TTCAACAGATGTGGAGATGGTACTAGAGGCGTGTGTTAGTGGGAGTATTGTTGGTAG 128
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 74 GTCAAAAGTTGGACTCGATGATCTCAGAGTCTTTTCCAGCCTTAATAATTCATGAAT 132
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 127 GTAGATGGTTGGATGATGCTCAGAGATCTTTTCCAACCTTAATGTTCTATGATT 69
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 15
CNS0100X
LOCUS
DEFINITION
  Drosophila melanogaster genome survey sequence sp6 end of BAC
  BACN03G04 of DrosBAC library from Drosophila melanogaster (fruit
  fly), genomic survey sequence.
ACCESSION
  AL098379
VERSION
  AL098379.1 GI:5609990
KEYWORDS
  GSS.
SOURCE
  fruit fly.
  Drosophila melanogaster
  Drosophila melanogaster
  Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
  Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
  Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
  1 (bases 1 to 1101)
AUTHORS
  Genoscope.
TITLE
  Direct Submission
JOURNAL
  Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
  BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr
  - Web : www.genoscope.cns.fr)
COMMENT
  Determination of this BAC-end sequence was carried out as part of a
  collaboration with the European Drosophila Genome Project (EDGP) -
  http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
  library (DrosBAC) was made by Alain Billaud at CEPH (Centre
  d'Etude du Polymorphisme Humain) with funding provided by a MRC
  project grant. The DNA was prepared from embryos by Alain Bucheton
  and Genevieve Payan. It has been constructed in the vector
  pBeloBAC11.
FEATURES
  Location/Qualifiers
    1..1101
      /organism="Drosophila melanogaster"
      /plasmid="pBeloBAC11"
      /db_xref="taxon:7227"
      /clone_lib="DrosBAC"
      /clone="BACN03G04"
      /note="end : SP6"
BASE COUNT
  195 a 108 c 131 g 161 t 506 others
ORIGIN
```

Query Match 7.6%; Score 55; DB 12; Length 1101;  
Best Local Similarity 19.4%; Pred. No. 22;  
Matches 102; Conservative 215; Mismatches 204; Indels 5; Gaps 2;

```
QY 154 TTTTTCAGCAAGTTGTTGGGATTTTAGTTTGGTTTCCCTGCTCACTGTTTCTTTCC 213
      :|:|: : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 428 KWTDRTRDDTKDTRDWDGKADTDWTDTKTWTWTWTGKKKWDTKTKKKWDDGG 487
      || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 214 TTGAACTGACTTTCATTTCCACATGAGAAATGCTGTAATTTCT---CAGTTACAAGTA 270
      || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 488 TTWTDTWKDKTDKTDKTDWADNAGWTDKDKTKAKKRTKTDTKKTRTGTGKK 547
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 271 GTCAATGGCTGCTAGAGAAACATTTAGGGAATACACTGAGTGAAGCAAC 330
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 548 TDRGMRKTDTWGWDWRADTRDGRADRAANAKDKADGRADTDATKDGWGTTRADTT 607
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 331 ACAGTGTACTGCCAACTAGCTTTGGGATTTGAGGCCACAGAGTTGTATATAAAT 390
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 608 DDTDADWMDKWDKWDADTKAGRKRGWGWGKKGKKTKTKKKKTDKTDKTDKTDK 667
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 391 TTGTTTAATGATATCTCGCCCTCGCCTTCATTAATGCTGTGTTTATGAACCACTCT 450
```

Search completed: August 2, 2002, 16:52:08  
Job time: 12209 sec

Mon Aug 5 11:51:35 2002

us-08-973-363-1.rst

Page 9

---

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 2, 2002, 13:35:09 ; Search time 1205.34 Seconds  
(without alignments)  
1029.858 Million cell updates/sec

Title: US-08-973-363-1

Perfect score: 723

Sequence: 1 CCCGTCGAGGTTTCAAGG.....CTTGAAACCTCCGACCGGG 723

Scoring table:

IDENTITY\_NUC  
Gap 10.0, Gapext 1.0

Searched:

1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N\_Geneseq\_032802.\*  
1: /SIDS1/gcgdata/hold-geneseq/geneq-emb1/NA1980.DAT.\*  
2: /SIDS1/gcgdata/hold-geneseq/geneq-emb1/NA1981.DAT.\*  
3: /SIDS1/gcgdata/hold-geneseq/geneq-emb1/NA1982.DAT.\*  
4: /SIDS1/gcgdata/hold-geneseq/geneq-emb1/NA1983.DAT.\*  
5: /SIDS1/gcgdata/hold-geneseq/geneq-emb1/NA1984.DAT.\*  
6: /SIDS1/gcgdata/hold-geneseq/geneq-emb1/NA1985.DAT.\*  
7: /SIDS1/gcgdata/hold-geneseq/geneq-emb1/NA1986.DAT.\*  
8: /SIDS1/gcgdata/hold-geneseq/geneq-emb1/NA1987.DAT.\*  
9: /SIDS1/gcgdata/hold-geneseq/geneq-emb1/NA1988.DAT.\*  
10: /SIDS1/gcgdata/hold-geneseq/geneq-emb1/NA1989.DAT.\*  
11: /SIDS1/gcgdata/hold-geneseq/geneq-emb1/NA1990.DAT.\*  
12: /SIDS1/gcgdata/hold-geneseq/geneq-emb1/NA1991.DAT.\*  
13: /SIDS1/gcgdata/hold-geneseq/geneq-emb1/NA1992.DAT.\*  
14: /SIDS1/gcgdata/hold-geneseq/geneq-emb1/NA1993.DAT.\*  
15: /SIDS1/gcgdata/hold-geneseq/geneq-emb1/NA1994.DAT.\*  
16: /SIDS1/gcgdata/hold-geneseq/geneq-emb1/NA1995.DAT.\*  
17: /SIDS1/gcgdata/hold-geneseq/geneq-emb1/NA1996.DAT.\*  
18: /SIDS1/gcgdata/hold-geneseq/geneq-emb1/NA1997.DAT.\*  
19: /SIDS1/gcgdata/hold-geneseq/geneq-emb1/NA1998.DAT.\*  
20: /SIDS1/gcgdata/hold-geneseq/geneq-emb1/NA1999.DAT.\*  
21: /SIDS1/gcgdata/hold-geneseq/geneq-emb1/NA2000.DAT.\*  
22: /SIDS1/gcgdata/hold-geneseq/geneq-emb1/NA2001A.DAT.\*  
23: /SIDS1/gcgdata/hold-geneseq/geneq-emb1/NA2001B.DAT.\*  
24: /SIDS1/gcgdata/hold-geneseq/geneq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	723	100.0	723	18	AA742755 Great tit CHD-W ge
2	54.8	7.6	12007	24	ABL32717 Human immune syste
3	51.2	7.1	18585	24	ABL34609 Human metastasis a
4	49.8	6.9	7669	22	AAS46597 Tumour suppressor
5	49.8	6.9	7669	24	ABL33846 Human immune syste
6	48.2	6.7	14147	22	AAS46744 Tumour suppressor
7	48	6.6	411	22	AA181186 Human polynucleoti
8	47.2	6.5	17738	24	ABL33538 Human immune syste
9	47	6.5	6419	24	ABL32267 Human immune syste

10	47	6.5	6861	24	ABL33228 Human immune syste
11	46.8	6.5	6109	24	ABL32326 Human immune syste
12	46.8	6.5	6109	24	AAS61077 Human gene regulat
13	46.8	6.5	6641	24	ABL32315 Human immune syste
14	46.8	6.5	7597	24	ABL33013 Human immune syste
15	46.6	6.4	7455	24	ABL33759 Human immune syste
16	46.4	6.4	5296	24	ABL33285 Human immune syste
17	46.2	6.4	6137	24	ABL33763 Human immune syste
18	46.2	6.4	6137	24	ABL34449 Human immune syste
19	46.2	6.4	17959	24	ABL32575 Human immune syste
20	46.2	6.4	73334	24	ABL34124 Human immune/haema
21	46	6.4	2205	22	AAK68629 Human immune syste
22	46	6.4	5204	24	ABL32901 Human immune syste
23	45.8	6.3	9963	24	ABL32694 Human immune syste
24	45.8	6.3	14006	24	ABL33958 Human immune syste
25	45.6	6.3	16167	24	ABL33083 Human immune syste
26	45.6	6.3	16167	24	ABL34529 Human metastasis a
27	45.6	6.3	40152	23	ABL17408 Drosophila melanog
28	45.4	6.3	395	22	AA181559 Human immune syste
29	45.4	6.3	1829	24	ABL34408 Human polynucleoti
30	45.2	6.3	393	22	AA181513 Human immune syste
31	45.2	6.3	5898	22	AAS45471 Chemically pretrea
32	45.2	6.3	5898	24	ABL33940 Human immune syste
33	45.2	6.3	5976	22	AAS46301 Tumour suppressor
34	45.2	6.3	15825	24	ABL33243 Human immune syste
35	45.2	6.3	15825	24	ABL34559 Human metastasis a
36	45	6.2	7635	24	ABL33431 Human immune syste
37	44.8	6.2	384	22	AA184685 Human polynucleoti
38	44.8	6.2	503	22	AA186964 Human polynucleoti
39	44.8	6.2	5576	24	ABL33516 Human immune syste
40	44.8	6.2	5610	22	AAS45462 Chemically pretrea
41	44.8	6.2	5610	24	AAD22332 Human immune syste
42	44.8	6.2	6127	24	ABL33614 Human immune syste
43	44.8	6.2	12359	24	ABL34047 Human immune syste
44	44.6	6.2	16994	24	ABL32989 Human immune syste
45	44.4	6.1	15872	22	AAS46520 Tumour suppressor

#### ALIGNMENTS

RESULT 1

AA742755 standard; DNA; 723 BP.

ID AA742755;

AC AA742755;

DT 12-MAR-1997 (first entry)

DE Great tit CHD-W gene probe (pGT-W insert).

XX Bird; sex determination; chromodomain-Helicase-DNA binding 1 Avian;

KW CHD-W; chromodomain-Helicase-DNA binding on the W chromosome; ss.

XX Parus major.

XX WO9639505-A1.

XX 12-DEC-1996.

XX 05-JUN-1996; 96WO-GB01341.

XX 06-JUN-1995; 95GB-0011439.

XX (ISIS-) ISIS INNOVATION LTD.

XX Griffiths R, Tiwari B;

XX WPI; 1997-043127/04.

XX Avian chromodomain-helicase-DNA binding genes determine sex in

PT birds - used for sex determin. and to control sex of progeny

XX

Claim 8; Fig 1; 76pp; English.

PS  
xx  
CC  
CC  
CC  
CC  
CC  
CC  
CC  
CC  
CC  
CC  
CC  
CC  
CC  
CC  
CC  
CC  
CC  
Sequence 723 BP; 188 A; 109 C; 164 G; 262 T; 0 other;

Claim 8; Fig 1; 76pp; English.

Probe GT-W (AAT42755) was cloned from a W chromosome specific PCR product derived from great tit. It represents part of an intron in a W-linked gene. By moving downstream from this sequence, a putative exon was isolated of a gene that was designated CHD-W due to its close identity to the mouse CHD-1 gene and its W location. This CHD-W fragment was used to isolate a similar gene (see also AAT42751) from chicken. This gene was not W-located, and was named CHD-1A. A second, W-located chicken gene (CHD-W) closely related to CHD-1A was also identified (see also AAT42754). The CHD-1A and CHD-W genes are useful for bird sex determin.

Query Match	100.0%;	Score 723;	DB 18;	Length 723;				
Best Local Similarity	100.0%;	Mismat. No. 1.3e-152;						
Matches 723;	Conservative 0;	Indels 0;	Gaps 0;					
QY	1	CCCGGTCGGAGGTTTC	AAAGGAATC	ACTAGATGTCGCACTTA	GTGCCATGGCTAGTTG	CAC	60	
DB	1	cccggtcggagggtttc	caaggaatgac	laga lgtg	gcacttagtgcca	tggtcctagctgac	60	
QY	61	AAGGTGATGTTGCT	AAAAAGTTG	CGACTCGATGATCT	CAGAGTATTTTTC	CAGGCTTAAT	120	
DB	61	aaggtgatggttggtc	aaaagttagct	gatgatc	cagagattttttt	cagacctaat	120	
QY	121	AATTCTATGAATTC	TGTAATTTAT	TCTTGATCTTTT	TGACGGAAGTTTG	TGGGGAATT	180	
DB	121	aattctatgaattctg	taatttattct	gatctctttt	gagcgaagttgtt	gttgggatt	180	
QY	181	TTAGTTTGGTTCCT	GTCACTGTTT	CTTCTCCTTG	AAACTGACTTTC	ATTGTCAACATG	240	
DB	181	ttagtttggtttccct	gtcactgtttt	cttcttccttg	gaactgaactt	catcttgcaacatg	240	
QY	241	AGAAATGCTGATT	TGTCAGGTT	TACAAGTAGT	GCAATGGCTGCT	TAGAAGTAGTGAGAAA	300	
DB	241	agaattgctgatttgc	aggtttaca	agtagtgc	aatgctctaga	agtagtgagaaa	300	
QY	301	CATTTAGGAAATAC	TGSGAGTGA	AGCAACACACT	GGTGACTGCCAA	ACTGTAGCTTTGGG	360	
DB	301	catctaggaaaatact	ggagtgaa	gcaaacacag	tggctactgc	caaacatgtagctttggg	360	
QY	361	ATTTGAGGAGCC	ACAGAGTTG	TATATAAATTT	TGTTTAAATGAT	TATCCTGCCCTGCCTTCC	420	
DB	361	atttgaggagccacag	agtgatata	aaattgtt	taatagatata	ctgcacctgcctcc	420	
QY	421	ATTAATTTGCTT	TTTTATGA	AACCACTCT	TTTTTTTTTTTT	TTTTTTTGGCTTCCTCA	480	
DB	421	attaattgctgtttt	atgaaccactc	tttttttttt	tttttttttt	tttttggcttcttca	480	
QY	481	TATCCTCTGGT	TAATGAGTT	TAATGCATTT	TAGAAGCACAT	GGCAGAACTAGGAGATCTGTGG	540	
DB	481	tatccttggttaa	tgagttaat	gcat ttag	aagcacat	tgccagaactagggagatctgtg	540	
QY	541	ATGACAGTGGT	PACAGGAGCT	CTGAAATTTT	TAGATAA	ACTATGAGAGTGGAAACAGAAAT	600	
DB	541	atgacagtgg	tacag	gagctctga	attttt	tagataaactatgagatg	ggaaacagaaa	600
QY	601	CTGAGGCTAGT	TTCTCTG	AGCTGACTGT	TAATTTTGT	GAGATATTTTCAAGACTACATTA	660	
DB	601	ctgagggtagtttct	gtagctgact	gtataatttt	gtgagaata	atttttcaagactacatta	660	
QY	661	GTTGTGTGTTT	TGAGGAAAA	TAAAAATG	TTTAAAGTTG	TCCATTCCTTGA	AACCTCCCGACC	720
DB	661	gttgtgtgtttg	agaaaaata	aaatgtt	taagttg	ccatttcc	ttgaaacctccgacc	720
QY	721	GGG	723					
DB	721	ggg	723					

## RESULT 2

ABL32717 standard; DNA; 12007 BP.  
 ABL32717;  
 26-MAR-2002 (first entry)  
 Human immune system associated gene SEQ ID NO: 690.  
 Human; immune system disease; cytosine methylation; antiasthmatic;  
 antiarteriosclerotic; anianaemic; cytosatic; nootropic;  
 neuroprotective; anti-Hiv; anticonvulsant; ophthalmological;  
 antirheumatic; antiarthritic; antidiabetic; antipsoriatic;  
 antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;  
 acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;  
 neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;  
 gene; ds.  
 Homo sapiens.  
 WO200200928-A2.  
 03-JAN-2002.  
 02-JUL-2001; 2001WO-EP07537.  
 30-JUN-2000; 2000DE-1032529.  
 01-SEP-2000; 2000DE-1043826.  
 (EPIC-) EPIGENOMICS AG.  
 Olek A, Piepenbrock C, Berlin K;  
 WPI; 2002-130909/17.  
 Nucleic acid comprising fragment of chemically modified gene, useful  
 for diagnosis and treatment of diseases associated with abnormal  
 cytosine methylation -  
 Claim 1; SEQ ID NO 690; 32pp + Sequence Listing; German.  
 The present invention provides a number of human immune system associated  
 genes which are modified by the methylation of cytosines. The sequences  
 can be used in the diagnosis and treatment of immune system disorders,  
 including eye diseases such as retinopathy, neovascular glaucoma and  
 macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
 leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
 rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
 diseases. The present sequence is a gene of the invention.  
 Sequence 12007 BP; 3266 A; 114 C; 2487 G; 6140 T; 0 other;

[illegible]









PR 01-SEP-2000; 2000DE-1043826.  
XX  
XX (EPIG-) EPIGENOMICS AG.  
PA  
XX  
XX Olek A, Piepenbrock C, Berlin K;  
PI  
XX WPI: 2002-130909/17.  
DR

The present invention provides a number of human immune system associated genes which are modified by the methylation of cytokines. The sequences can be used in the diagnosis and treatment of immune system disorders including eye diseases such as retinopathy, mis, glaucoma and macular degeneration. The present invention also provides a number of human immune system associated genes which are modified by the methylation of cytokines. The sequences can be used in the diagnosis and treatment of immune system disorders including eye diseases such as retinopathy, mis, glaucoma and macular degeneration. The present invention also provides a number of human immune system associated genes which are modified by the methylation of cytokines. The sequences can be used in the diagnosis and treatment of immune system disorders including eye diseases such as retinopathy, mis, glaucoma and macular degeneration.

CC genes which are modified by the methylation of cytosines. The sequence  
CC can be used in the diagnosis and treatment of immune system disorders  
CC including eye diseases such as retinopathy, neovascular glaucoma and  
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
CC diseases. The present sequence is a gene of the invention.  
XX  
SQ Sequence 6861 BP; 1740 A; 246 C; 1607 G; 3268 T; 0 other;

Query Match 6.5%; Score 47; DB 24; Length 6861;  
Best local Similarity 50.7%; Pred. No. 0.33;  
Matches 113; Conservative 0; Mismatches 110; Indels 0; Gaps

QY 356 TTGGGATTTCGAGGAGCCACAGAGTGTGATATAAATTTGTTTAATGATCTCCCTGC 41  
Db 417 ttaggaaattggataaaaaatatataaaatttatataatatataattattattattatta 47  
QY 416 CTCCATTAAATGCTGTTTGTATGAACCACTCTTTTTTTTTTTTTTTTTTTTGGCTT 47  
Db 477 acgaggttgagattttttttatatgtttattgtgaattgtatttttttttttgtagttt 53  
QY 476 CTCTATPCTCTGGTGAATGAGTGTAAATGCATTTAGAAAGCACATGGCAGAACTAGGAGATC 53  
Db 537 attatttttggttgatatatttttttatagtaaaagaagaagaaatgatttatgttt 59  
QY 536 TGTGATGACAGTGGTACAGGAGCTCTGAATTTTTTAGATAAA 578  
Db 597 tatatagctatttggagagatagtttatgaaaatttgttaagaaa 639

597 tatatagtattttggagatagttatgaanaattttaagaaa 639

Db	RESULT	11	
	ABL32326		
	ID	ABL32326 standard; DNA; 6109 BP.	
	XX		
	AC		
	XX	ABL32326;	
	DT	26-MAR-2002 (first entry)	
	XX		
	DE	Human immune system associated gene SEQ ID NO: 299.	
	XX		
	KW	Human: immune system disease; cytosine methylation; antiasthmatic;	
	KW	antiartherosclerotic; antianaemic; cyostatic; nootropic;	
	KW	neuroprotective; anti-HIV; anticonvulsant; ophthalmological;	
	KW	antirheumatic; antiarthritis; antidiabetic; antipsoriatic;	
	KW	antinflammatory; cancer; eye disease; arteriosclerosis; anaemia;	
	KW	acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;	
	KW	neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;	
	KW	gene; ds.	
	XX		
	OS	Homo sapiens.	
	XX		
	PN	WO200200928-A2.	
	XX		
	PD	03-JAN-2002.	
	XX		
	DE	02-JUN-2001. 2001WO-EP07537	

597 tatatagtattttggagatagttatgaanaattttaagaaa 639

Db	RESULT	11	
	ABL32326		
	ID	ABL32326 standard; DNA; 6109 BP.	
	XX		
	AC		
	XX	ABL32326;	
	DT	26-MAR-2002 (first entry)	
	XX		
	DE	Human immune system associated gene SEQ ID NO: 299.	
	XX		
	KW	Human: immune system disease; cytosine methylation; antiasthmatic;	
	KW	antiartherosclerotic; antianaemic; cyostatic; nootropic;	
	KW	neuroprotective; anti-HIV; anticonvulsant; ophthalmological;	
	KW	antirheumatic; antiarthritis; antidiabetic; antipsoriatic;	
	KW	antinflammatory; cancer; eye disease; arteriosclerosis; anaemia;	
	KW	acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;	
	KW	neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;	
	KW	gene; ds.	
	XX		
	OS	Homo sapiens.	
	XX		
	PN	WO200200928-A2.	
	XX		
	PD	03-JAN-2002.	
	XX		
	DE	02-JUN-2001. 2001WO-EP07537	

XX 30-JUN-2000; 2000DE-1032529.  
PR 01-SEP-2000; 2000DE-1043826.  
XX (EPIG-) EPIGENOMICS AG.  
XX Olek A, Piepenbrock C, Berlin K;  
PI WPI; 2002-130909/17.  
XX Nucleic acid comprising fragment of chemically modified gene, useful  
PT for diagnosis and treatment of diseases associated with abnormal  
PT cytosine methylation -  
XX Claim 1; SEQ ID NO 299; 32pp + Sequence Listing; German.  
XX The present invention provides a number of human immune system associated  
XX genes which are modified by the methylation of cytosines. The sequences  
CC can be used in the diagnosis and treatment of immune system disorders,  
CC including eye diseases such as retinopathy, neovascular glaucoma and  
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
CC diseases. The present sequence is a gene of the invention.  
XX  
SQ Sequence 6109 BP; 1232 A; 195 C; 1734 G; 2944 T; 4 other;  
  
Query Match 6.5%; Score 46.8; DB 24; Length 6109;  
Best Local Similarity 45.8%; Pred. No. 0.36;  
Matches 162; Conservative 0; Mismatches 192; Indels 0; Gaps 0;  
  
QY 355 TTGTGGGATTTGAGGAGCCACAGAGTTGTATATAAATTTGTTTAATGATATCCCTGCCCTG 414  
Db 1860 ttattattattattattattattattattattattattattattattattattattatttt 1919  
  
QY 415 CCTTCCATTAATGCTGTTTATGAACACACTCTTTTTTTTTTTTTTTTTTTTGGCT 474  
Db 1920 ttattattattattattattattattattattattattattattattattattatttt 1979  
  
QY 475 TCTTCATATCCTGCTGTAATGAGTTAATGATTAAGTATGATATCCCTGCCCTG 534  
Db 1980 tctttattattattattattattattattattattattattattattattattattatttt 2039  
  
QY 535 CTGTGGATGACAGTGTGACAGGAGCTGTAATTTTATAGATAACATGAGAGTGAAC 594  
Db 2040 ttattattattattattattattattattattattattattattattattattatttt 2099  
  
QY 595 AGAAATCTGAGGCTAGTTCTGAGCTGACTGTAATTTTGTGAGATAATTTCAAGACT 654  
Db 2100 ttattattattattattattattattattattattattattattattattattatttt 2159  
  
QY 655 ACATTAGTCTGTTGAGGAAATAAATGTTTAAGTTGTCCTTGA 708  
Db 2160 tattattattattattattattattattattattattattattattattattattttta 2213  
  
RESULT 12  
ID AAS61077  
XX AAS61077 standard; DNA; 6109 BP.  
XX AAS61077;  
XX 29-JAN-2002 (first entry)  
XX Human gene regulation-associated gene oligonucleotide #32.  
XX Human; Gene regulation-associated gene; severe combined immunodeficiency;  
KW cardiac damage; inflammatory response; Haemophilia; Werner syndrome;  
KW asthma; HDR syndrome; congenital heart defect; Saethre-Chotzen syndrome;  
KW renal disease; Preclampsia; cardiac allograft vascular disease;  
KW colorectal cancer; thyroid cancer; oesophageal cancer; ds; tumour;  
KW immunostimulant; cardiant; antiinflammatory; coagulant; antiasthmatic;

KW nephrotropic; gynecological; anti-tumour; immunosuppressive; cytostatic.  
XX Homo sapiens.  
XX WQ200177375-A2.  
XX 18-OCT-2001.  
XX 06-APR-2001; 2001WO-EP03968.  
XX 06-APR-2000; 2000DE-1019058.  
PR 07-APR-2000; 2000DE-1019173.  
PR 30-JUN-2000; 2000DE-1032529.  
PR 01-SEP-2000; 2000DE-1043826.  
XX (EPIG-) EPIGENOMICS AG.  
XX Olek A, Piepenbrock C, Berlin K;  
PI WPI; 2002-017470/02.  
XX New nucleic acid sequences from chemically modified genes associated  
PT with gene regulation, useful for analysing cytosine methylations for  
PT diagnosis and therapy of diseases e.g. severe combined immunodeficiency  
PT disease -  
XX Claim 1; SEQ ID NO 33; 26pp; English.  
XX The invention relates to 224 nucleic acid sequences comprising at least  
CC 18 bases of a chemically pretreated gene associated with gene regulation  
CC selected from 43 known genes (or complementary sequences). The  
CC chemical pretreatment converts cytosine bases unmethylated at the  
CC 5-position to uracil or another base with hybridisation behaviour  
CC dissimilar to cytosine, to enable analysis of cytosine methylations.  
CC The DNA sequences, oligomers (or sets/arrays) and method are  
CC useful in the diagnosis of diseases (or predisposition to diseases)  
CC associated with gene regulation and in therapy of such diseases, by  
CC enabling analysis of the cytosine methylation patterns of such genes,  
CC kits are provided. They are especially useful in diagnosis  
CC and therapy of e.g. severe combined immunodeficiency disease, cardiac  
CC disorders, haemophilia, solid tumours and cancer, Werner syndrome,  
CC asthma, HDR syndrome, Saethre-Chotzen syndrome, renal disease,  
CC preclampsia, graft versus-host disease. The present sequence is a  
CC sequence included in the sequence data for this specification and is  
CC associated with the human gene regulation-associated genes.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 6109 BP; 1232 A; 195 C; 1734 G; 2944 T; 4 other;  
  
Query Match 6.5%; Score 46.8; DB 24; Length 6109;  
Best Local Similarity 45.8%; Pred. No. 0.36;  
Matches 162; Conservative 0; Mismatches 192; Indels 0; Gaps 0;  
  
QY 355 TTGTGGGATTTGAGGAGCCACAGAGTTGTATATAAATTTGTTTAATGATATCCCTGCCCTG 414  
Db 1860 ttattattattattattattattattattattattattattattattattattattttt 1919  
  
QY 415 CCTTCCATTAATGCTGTTTATGAACACACTCTTTTTTTTTTTTTTTTTTTTGGCT 474  
Db 1920 ttattattattattattattattattattattattattattattattattattatttt 1979  
  
QY 475 TCTTCATATCCTGCTGTAATGAGTTAATGATTAAGTATGATATCCCTGCCCTG 534  
Db 1980 tctttattattattattattattattattattattattattattattattattattatttt 2039  
  
QY 535 CTGTGGATGACAGTGTGACAGGAGCTGTAATTTTATAGATAACATGAGAGTGAAC 594  
Db 2040 ttattattattattattattattattattattattattattattattattattatttt 2099



163 GAAGTTTGTTGGGGATTTTAGTTGGTTCCCTGTCAC TGTTTCTTTCTTCTTGAAACTG 222

The present invention provides a number of human immune system associated genes which are modified by the methylation of cytosines. The sequences can be used in the diagnosis and treatment of immune system disorders, including eye diseases such as retinopathy, neovascular glaucoma and macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis, rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel diseases. The present sequence is a gene of the invention.

```
XX
SQ Sequence 7597 BP; 1965 A; 109 C; 1790 G; 3733 T; 0 other;

Query Match      6.5%; Score 46.8; DB 24; Length 7597;
Best Local Similarity 45.8%; Pred. No. 0.37;
Matches 162; Conservative 0; Mismatches 192; Indels 0; Gaps 0;

QY 355 TTGTGGATTTGAGGCCACAGAGTTGTATATAAATTTGTTAATGATATCCTGCCCCG 414
Db 473 ttatgattattattattatgattattattattattattattattattattattatt 532
QY 415 CCTTCATTAATGCTGTGTTTATGAACACACCTCTTTTTTTTTTTTTTTTTTTGGCT 474
Db 533 ttatttattattattattattattattattattattattattattattattattatt 592
QY 475 TCTTCATATCCTGTGGTAATGAGTTAATGCATTTAGACACATGCGACAACTAGAGAT 534
Db 593 ttatttattattattattattattattattattattattattattattattattatt 652
QY 535 CTCTGGATGACAGCTGTACAGGAGCTGTAATTTTATGATAAATGATGAGAGTGAAC 594
Db 653 gtgttattattattattattattattattattattattattattattattattattat 712
QY 595 AGAAATCTGAGGCTAGTTTCTTGAGCTGACTGTAAATTTTGTGAGAAATATTTTCAAGACT 654
Db 713 ttatatattattattattattattattattattattattattattattattattatt 772
QY 655 ACATTAGTGTGTGTTGAGGAAAATAAATGTTTAAAGTTGTCATTCCTTGA 708
Db 773 ttatttattattattattattattattattattattattattattattattattat 826

RESULT 15
ABL33759
ID ABL33759 standard; DNA; 7455 BP.
AC ABL33759;
DT 26-MAR-2002 (first entry)
DE Human immune system associated gene SEQ ID NO: 1732.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytosolic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritis; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW gene; ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP07537.
XX
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
PA (EPIC-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
XX Nucleic acid comprising fragment of chemically modified gene, useful
XX for diagnosis and treatment of diseases associated with abnormal
XX cytosine methylation
```

PS Claim 1; SEQ ID NO 1732; 32pp + Sequence Listing; German.

XX The present invention provides a number of human immune system associated genes which are modified by the methylation of cytosines. The sequences can be used in the diagnosis and treatment of immune system disorders, including eye diseases such as retinopathy, neovascular glaucoma and macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis, rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel diseases. The present sequence is a gene of the invention.

XX

SQ Sequence 7455 BP; 2550 A; 38 C; 1350 G; 3517 T; 0 other;

Query Match 6.4%; Score 46.6; DB 24; Length 7455;

Best Local Similarity 48.3%; Pred. No. 0.41;

Matches 159; Conservative 0; Mismatches 169; Indels 1; Gaps 1;

QY 379 TTGTATATAAATTTGTTTAAATGATATCCTGCCCTGCCCTCCATTAATTTGCTTTGTTTAT 438

Db 4345 ttaaaaataaattgttcgattatttggggaagttaaaatgattttttttttttttttt 4404

QY 439 GAAACACACTCTTTTTTTTTTTTTTTTTTTTTTTTGGCTTCTTCATATCCTGTGGTAATGAGT 498

Db 4405 ttt 4464

QY 499 TAATGCATTTAGAAAGCACATGCGACAGTACAGAGATCTGTGGATGACAGTGG-TACAGGA 557

Db 4465 taataatattaggattgttttttagtagatttaaggcattttattgttttagtaataaag 4524

QY 558 GCTCTGAATTTTATGATAAACTATGAGAGTGGAAACAGAAATCTGAGGCTAGTTTCTTG 617

Db 4525 taaatattagtttaaaaggattataataatagtaaatattatgtgtaagaataattgat 4584

QY 618 AGCTGACTGTAAATTTTGTGAGAAATATTTTCAAGACTACATTAATTTGCTTTGAGGAA 677

Db 4585 attttaatatatatatttttaaaattagtttaaaatttttttttttttttttttttt 4644

QY 678 AAATAAAATGTTTAAAGTTGTCATTCCTT 706

Db 4645 aatattaatattaggattattatttttttttttttttttttttttttttttttt 4673

Search completed: August 2, 2002, 20:44:37

Job time: 25768 sec

Mon Aug 5 11:51:34 2002

us-08-973-363-1.rng

Page 11

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: August 2, 2002, 20:28:20 ; Search time 12674 Seconds  
(without alignments)  
252.625 Million cell updates/sec

Title: US-08-973-363-2  
Perfect score: 153  
Sequence: 1 ATCTTCAGATGATCCTGA.....CTCAGAGACTTTGTGGTGG 153

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues  
Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*  
1: gb\_ba.\*  
2: gb\_htg.\*  
3: gb\_in.\*  
4: gb\_ov.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_sts.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vi.\*  
15: em\_ba.\*  
16: em\_fun.\*  
17: em\_hum.\*  
18: em\_in.\*  
19: em\_mu.\*  
20: em\_mu.\*  
21: em\_or.\*  
22: em\_ov.\*  
23: em\_pat.\*  
24: em\_ph.\*  
25: em\_pl.\*  
26: em\_ro.\*  
27: em\_sts.\*  
28: em\_un.\*  
29: em\_vi.\*  
30: em\_htg\_hum.\*  
31: em\_htg\_inv.\*  
32: em\_htg\_other.\*  
33: em\_htg\_inv.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
------------	-------	-------------	--------	-------	-------------

1	153	100.0	153	6	A58683	A58683 Sequence 2
2	105.6	69.0	153	6	A58685	A58685 Sequence 4
3	105.6	69.0	153	6	A58686	A58686 Sequence 5
4	102.4	66.9	153	6	A58684	A58684 Sequence 3
5	102	66.9	5349	10	MUSCHD1X	L10410 Mouse DNA-b
6	77	50.3	5947	9	AF006513	AF006513 Homo sapi
7	77	50.3	101220	9	AC092372	AC092372 Homo sapi
8	77	50.3	134365	9	AC012624	AC012624 Homo sapi
9	77	50.3	143079	2	AC021449	AC021449 Homo sapi
10	77	50.3	145659	2	AC008531	AC008531 Homo sapi
11	77	50.3	193446	2	AC091946	AC091946 Homo sapi
12	77	50.3	219258	9	AC022121	AC022121 Homo sapi
13	77	50.3	276181	2	AC092382	AC092382 Homo sapi
14	68.4	44.7	6608	6	A58691	A58691 Sequence 10
15	68.4	44.7	6872	5	AF004397	AF004397 Gallus ga
16	67.4	44.1	1316	6	A58696	A58696 Sequence 15
17	54	35.3	1311	6	AR029026	AR029026 Sequence
18	52.2	34.1	2292	5	D14316	D14316 delta-cryst
19	41.6	27.2	2348	8	AY056298	AY056298 Arabidops
20	41.6	27.2	88010	8	ATT14D3	AL138649 Arabidops
21	41.6	27.2	89214	8	ATT14D17	AL335932 Arabidops
22	41.2	26.9	446	6	AX333515	AX333515 Sequence
23	41.2	26.9	7764	9	AF006514	AF006514 Homo sapi
24	41.2	26.9	177951	2	AC074383	AC074383 Homo sapi
25	41.2	26.9	183858	9	AC013394	AC013394 Homo sapi
26	41	26.8	93893	2	AC087641	AC087641 Homo sapi
27	40.8	26.7	172008	2	AC091344	AC091344 Rattus no
28	39.8	25.6	201410	9	HSP12	AC074041 Rattus no
29	39.2	25.6	201410	10	AC074041	AL133500 Homo sapi
30	37.8	24.7	111845	2	AC097570	AC097570 Rattus no
31	37.4	24.4	1537	9	HSB02032	AL137347 Homo sapi
32	37.4	24.4	6118	6	AX345933	AX345933 Sequence
33	37.4	24.4	80730	10	AF332579	AF332579 Mus muscu
34	37.4	24.4	175944	2	AC026878	AC026878 Mus muscu
35	37.4	24.4	250227	2	AC093363	AC093363 Mus muscu
36	36.6	23.9	113587	9	AL590702	AL590702 Human DNA
37	36.6	23.9	194265	2	AC094841	AC094841 Rattus no
38	36.6	23.9	220216	9	AC021019	AC021019 Homo sapi
39	36.4	23.8	160763	2	AC098534	AC098534 Rattus no
40	36.2	23.7	137129	2	HSAC001236	AC001236 Homo sapi
41	36.2	23.7	176282	2	AC068507	AC068507 Homo sapi
42	36.2	23.7	179141	2	AC019091	AC019091 Homo sapi
43	36.2	23.7	185074	2	AC093526	AC093526 Homo sapi
44	36.2	23.7	191911	9	AC098965	AC098965 Homo sapi
45	36.2	23.7	199144	2	AC105129	AC105129 Homo sapi

ALIGNMENTS

RESULT	1	A58683	Sequence 2 from Patent WO9639505.	153 bp	DNA	linear	PAT 06-MAR-1998
LOCUS	A58683	Sequence 2 from Patent WO9639505.					
DEFINITION	A58683	Sequence 2 from Patent WO9639505.					
ACCESSION	A58683	Sequence 2 from Patent WO9639505.					
VERSION	A58683.1	GI:3714246					
KEYWORDS		unidentified.					
SOURCE		unclassified.					
ORGANISM		unclassified.					
REFERENCE		1 (bases 1 to 153)					
AUTHORS		Griffiths, R. and Tiwari, B.					
TITLE		AVIAN GHD GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN BIRDS					
JOURNAL		Patent: WO 9639505-A 2 12-DEC-1996;					
COMMENT		ISS INNOVATION (GB)					
FEATURES		Other publication AU 590696 961224.					
		Location/Qualifiers					
		1..153					
		/organism="unidentified"					
		/db_xref="taxon:32644"					
BASE COUNT		61 a 37 c 27 g 28 t					
ORIGIN							



MUSCHD1X  
LOCUS 5349 bp mRNA linear ROD 19-SEP-1996  
DEFINITION Mouse DNA-binding protein (CHD-1) mRNA, complete cds.  
ACCESSION L10410 X66028  
VERSION L10410.1 GI:455014  
KEYWORDS DNA binding protein.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
Delmas,V., Stokes,D.G. and Perry,R.P.  
A mammalian DNA-binding protein that contains a chromodomain and an  
SNF2/SWI2-like helicase domain  
Proc. Natl. Acad. Sci. U.S.A. 90 (6), 2414-2418 (1993)  
93211972  
REFERENCE 1 (bases 1 to 5349)  
AUTHORS Perry,R.P.  
DIRECT SUBMISSION  
JOURNAL Submitted (08-APR-1993) Robert Perry, The Fox Chase Cancer Center,  
7701 Burholme Avenue, Philadelphia, PA 19111, USA  
COMMENT On Feb 16, 1994 this sequence version replaced gi:293322.  
FEATURES  
Location/Qualifiers  
1..5349  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/cell\_type="plasmacytoma"  
1..5349  
/gene="CHD-1"  
171..5306  
/gene="CHD-1"  
/codon\_start=1  
/product="DNA-binding protein"  
/protein\_id="AAB08486.1"  
/db\_xref="GI:455015"  
/translation="MNGHSDSESVNRSGSSQSGDDCGSAGSGSGSSGSSDSSGSS  
SQSGSDSDSGSSQSGSESDTSRENKVOAKPKVDGAEPKWSKSPSILAVORSAML  
RKQQAQOORPASSNSGSESDSSSDSSGAKRKHNDQWQSGSPSOLG  
SDSESEERKSSCGDTESDYEPKVRSPQRKSKNGKILGOKRQIDSEDE  
DDEYDNDKRSRQATVNVSYKDEDEMTSDLLCEGVEDVPQDEEETIERVM  
DCVRGKATGATTITAVADGDPNAGFNKPEGOIYLIKWKGWSHINTWETE  
TLKQVNRGKMLDNYKKQETKRLKNASPEDVEYVNCQELTDLHKGQIVERI  
IAISNOKSAAGLPDYCKWQGLPYSGHCEGLRDYQLNGLWLAHSCWKNCSILADEML  
DKVLKORPRFVALKKQPSYIGGHEGLRDYQLNGLWLAHSCWKNCSILADEML  
GKTQITISFLNLPHEHOLYGPFLVPLSTLTSQREIQTWASQNNVAVLGDINSR  
NMLRTHWHPOTKRLFNLTITPLNSKELSLHFTMPKPSWDEPEEHRKNDLSL  
YLKLEPELPRVKKQVEKSLPAKVEQILLEMESALQOYKWLITRYKALSKGSK  
GSTSGFLNLMELKCCNHCYLIKPPDNNEFYKQEAOLHILRSSGKLLILDLKLLI  
RERGNVLIFSOMVRLDILAEYLYKQPPQRLDGSIGKELKQALDHFNAESDF  
CFYLLSTRAGGLINLASADTVIFSDMNPQNDLQAQRAHRIQKQVNIYRLVYK  
SVEDLIERAKKMWLQVIRMDTGTLYHTGSAPSSSTPFNKEALSAILKFGAE  
ELKPEGEQEPQEMDIDELKRAETHENEPGLSVGDELLSQFVANFSNMDEDDI  
ELFPERNKWEEIIPEDQRRLKEERKQLEETIYMLPRMNCARQISFNGSGRRS  
RSRYSGSDSDSISERPKRRKRPRTIPRENIGFSDAEIRRFIKSYKFKGGLERL  
DATAARDELVDKSEITLRLGLVHNGCIKALKDSSTGTERAGRLGVKQGTFRISG  
VQNAKLIVTAHEDELIPLHKSIPDPERKQYTPCHTAAHFDIDWKEDDSSNLLIG  
IYEGYGSWEMIKMDPDLSTHLKLPDPDKPOAKOLOTRADYLIKLLSRLAKREA  
ORLCAGGSKRRKTRAKSKAMKIKVEEIKSDSPSEKSEDEDDDKLNDKSPESK  
DRKSKVSDAPVHTASGEPVPIAESELDQKTFICKERNRPVKAALQDLRPEK  
GLSERQLEHTRQCLIKIGDHTECLKEYSNPQIKQWRKLNWIFVSKFTEDARKLH  
KLYKHAIKRQESQNSDQNSVATVYIRNPDMERLKNENTHSDSDSDSDSDHLH  
QYHDHDKRHQDSYKSKSRKRPYVFSFNGKDRWDHYRQDSRYSDREKHKRLDD  
HRSREHPSLEGLKDRCHSDHSHSDHRSSEHTHHKSRDRIYLSWDQID  
HRAASSGPRPLDQSRPYGSRKSPFHSABHRSTPEHTWSSRK"

BASE COUNT 1739 a 1067 c 1319 g 1224 t  
ORIGIN

QY 52 AAAAAACCAACAAACAGTTACAGACCGTGCAGACTACCTCATCAAACTACTTAGC 111  
|||||  
Db 4047 AAAAAACCAACAAACAGTTACAGACCGTGCAGACTACCTCATCAAACTACTTAGC 4106  
|||||  
QY 112 AGAGATCTTGC AAAAGAGAGGCTCAGAGACTTTGTGGTGGC 153  
|||||  
Db 4107 AGAGATCTTGC AAAAGAGAGGCTCAGAGACTTTGTGGTGGC 4148  
|||||  
RESULT 6  
AF006513  
LOCUS Homo sapiens CHD1 mRNA, complete cds.  
DEFINITION AF006513  
ACCESSION AF006513  
VERSION AF006513.1 GI:2645428  
KEYWORDS human.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 5947)  
AUTHORS Woodage,T., Basrai,M.A., Baxevasis,A.D., Hietler,P. and Collins,F.S.  
TITLE Characterization of the CHD family of proteins  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 94 (21), 11472-11477 (1997)  
MEDLINE 97470591  
REFERENCE 2 (bases 1 to 5947)  
AUTHORS Woodage,T.  
TITLE Direct Submission  
JOURNAL Submitted (03-JUN-1997) Laboratory of Gene Transfer, National Human  
Genome Research Institute, National Institutes of Health, 49  
Convent Drive, Bethesda, MD 20892-4442, USA  
FEATURES  
Location/Qualifiers  
1..5947  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="5"  
/map="5q15-21; near WI-5811"  
1..5947  
/gene="CHD1"  
164..5293  
/gene="CHD1"  
/codon\_start=1  
/product="CHD1"  
/protein\_id="AAB87381.1"  
/db\_xref="GI:2645429"  
/translation="MNGHSDSESVNRSGSSQSGDDCGSAGSGSGSSGSSDSSGSS  
SQSGSDSDSGSSQSGSESDTSRENKVOAKPKVDGAEPKWSKSPSILAVORSAIL  
KKQOQOQOQOHOASNSGSESDSSSDSSSVKPKKHKDDQWQSGSPSOLG  
GDSSESEERKSSCGDTESDYEPKVRSPQRKSKNGKILGOKRQIDSEDE  
DDEYDNDKRSRQATVNVSYKDEDEMTSDLLCEGVEDVPQDEEETIER  
FMDRIGKATGATTITAVADGDPNAGFNKPEGOIYLIKWKGWSHINTWETE  
EETLKQVNRGKMLDNYKKQETKRLKNASPEDVEYVNCQELTDLHKGQIVRI  
IAISNOKSAAGLPDYCKWQGLPYSGHCEGLRDYQLNGLWLAHSCWKNCSILADEM  
FKDVKLORPRFVALKKQPSYIGGHEGLRDYQLNGLWLAHSCWKNCSILADEM  
GKTQITISFLNLPHEHOLYGPFLVPLSTLTSQREIQTWASQNNVAVLGDINSR  
NMLRTHWHPOTKRLFNLTITPLNSKELSLHFTMPKPSWDEPEEHRKNDLSL  
YLKLEPELPRVKKQVEKSLPAKVEQILLEMESALQOYKWLITRYKALSKGSK  
GSTSGFLNLMELKCCNHCYLIKPPDNNEFYKQEAOLHILRSSGKLLILDLKLLI  
RERGNVLIFSOMVRLDILAEYLYKQPPQRLDGSIGKELKQALDHFNAESDF  
CFYLLSTRAGGLINLASADTVIFSDMNPQNDLQAQRAHRIQKQVNIYRLVYK  
SVEDLIERAKKMWLQVIRMDTGTLYHTGSAPSSSTPFNKEALSAILKFGAE  
ELKPEGEQEPQEMDIDELKRAETHENEPGLSVGDELLSQFVANFSNMDEDDI  
ELFPERNKWEEIIPEDQRRLKEERKQLEETIYMLPRMNCARQISFNGSGRRS  
RSRYSGSDSDSISERPKRRKRPRTIPRENIGFSDAEIRRFIKSYKFKGGLERL  
DATAARDELVDKSEITLRLGLVHNGCIKALKDSSTGTERAGRLGVKQGTFRISG  
VQNAKLIVTAHEDELIPLHKSIPDPERKQYTPCHTAAHFDIDWKEDDSSNLLIG  
IYEGYGSWEMIKMDPDLSTHLKLPDPDKPOAKOLOTRADYLIKLLSRLAKREA  
ORLCAGGSKRRKTRAKSKAMKIKVEEIKSDSPSEKSEDEDDDKLNDKSPESK  
DRKSKVSDAPVHTASGEPVPIAESELDQKTFICKERNRPVKAALQDLRPEK  
GLSERQLEHTRQCLIKIGDHTECLKEYSNPQIKQWRKLNWIFVSKFTEDARKLH  
KLYKHAIKRQESQNSDQNSVATVYIRNPDMERLKNENTHSDSDSDSDHLH  
QYHDHDKRHQDSYKSKSRKRPYVFSFNGKDRWDHYRQDSRYSDREKHKRLDD  
HRSREHPSLEGLKDRCHSDHSHSDHRSSEHTHHKSRDRIYLSWDQID  
HRAASSGPRPLDQSRPYGSRKSPFHSABHRSTPEHTWSSRK"

Query Match 66.7%; Score 102; DB 10; Length 5349;  
Best Local Similarity 100.0%; Pred. No. 3.2e-18;  
Matches 102; Conservative 0; Mismatches 0; Indels 0; Gaps 0;





JOURNAL  
REFERENCE  
AUTHORS

Unpublished  
2 (bases 1 to 143079)  
Birren B., Linton L., Nusbaum C., Lander E., Abraham H., Allen N.,  
Anderson S., Balakrishnan J., Barna N., Beckerly R., Beda F.,  
Bohlander L., Boukhgalter S., Brown A., Burkett G., Castle A.,  
Chorel Y., Colangelo M., Collins S., Collymore A., Cooke P.,  
Dekrelans K., Dewar K., Domino M., Doyle M., Fenestor J.,  
Ferreira P., Fitchugh W., Forrest C., Gage D., Galagan J.,  
Gardner S., Grant G., Hago B., Heaford A., Horton L.,  
Howland J.C., Johnson R., Jones C., Kann L., Karatas A., Klein J.,  
Lander E., Lehotzky J., Levine R., Liu C., Liu G., Locke K.,  
Macdonald P., Marquis N., McEwan P., McGurk A., McKernan K.,  
McPheters R., Meldrum J., Menes L., Morrow J., Naylor J.,  
Norman C.H., O'Connor T., O'Donnell P., Oliver T.M., Peterson K.,  
Pierre N., Pisanl C., Pollara V., Raymond C., Riley R., Rothman D.,  
Roy A., Santos R., Severy P., Spencer B., Stange-Thomann N.,  
Stojanovic N., Subramanian A., Talamas J., Tesfaye S., Theodore J.,  
Tirrell A., Vassiliev H., Viel R., Vo A., Wu X., Wyman D., Ye W.J.,  
Zimmer A. and Zody M.

TITLE  
JOURNAL

Direct Submission  
Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Sep 10, 2000 this sequence version replaced gi:7407963.  
All repeats were identified using RepeatMasker.

## COMMENT

Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>  
-----  
Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
-----  
Project Information  
Center project name: L3154  
Center clone name: 58\_ML12  
-----  
Summary Statistics  
Sequencing vector: M13; M77815; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 134743 bases at least Q40  
Consensus quality: 139227 bases at least Q30  
Consensus quality: 140814 bases at least Q20  
Insert size: 144000; agarose-fp  
Insert size: 142179; sum-of-contigs  
Quality coverage: 4.6 in Q20 bases; agarose-fp  
Quality coverage: 4.7 in Q20 bases; sum-of-contigs  
-----

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 10 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 38820: contig of 38820 bp in length  
\* 38921 38920: gap of 100 bp  
\* 38921 40411: contig of 1491 bp in length  
\* 40412 40511: gap of 100 bp  
\* 40512 43279: contig of 2768 bp in length  
\* 43280 43379: gap of 100 bp  
\* 43380 46905: contig of 3526 bp in length  
\* 46906 47005: gap of 100 bp  
\* 47006 51830: contig of 4825 bp in length  
\* 51831 51930: gap of 100 bp  
\* 51931 62619: contig of 10689 bp in length  
\* 62620 62719: gap of 100 bp  
\* 62720 75408: contig of 12689 bp in length  
\* 75409 75508: gap of 100 bp  
\* 75509 92516: contig of 17008 bp in length  
\* 92517 92616: gap of 100 bp  
\* 92617 106409: contig of 13793 bp in length  
\* 106410 106509: gap of 100 bp  
\* 106510 143079: contig of 36570 bp in length.

FEATURES  
source

Location/Qualifiers  
1..143079  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_RP11="58M12"  
/clone\_lib="RPC1-11 Human Male BAC"

## misc\_feature

1..38820  
/note="assembly\_fragment"  
vector\_end:SP6  
vector\_side:left"

## misc\_feature

38921..40411  
/note="assembly\_fragment"

## misc\_feature

40512..43279  
/note="assembly\_fragment"

## misc\_feature

43380..46905  
/note="assembly\_fragment"

## misc\_feature

47006..51830  
/note="assembly\_fragment"

## misc\_feature

51931..62619  
/note="assembly\_fragment"

## misc\_feature

62720..75408  
/note="assembly\_fragment"

## misc\_feature

75509..92516  
/note="assembly\_fragment"

## misc\_feature

92617..106409  
/note="assembly\_fragment"

## misc\_feature

106510..143079  
/note="assembly\_fragment"  
clone\_end:T7  
vector\_side:right"

BASE COUNT 43971 a 26246 c 26678 g 45278 t 906 others  
ORIGIN

Query Match 50.3%; Score 77; DB 2; Length 143079;  
Best Local Similarity 94.1%; Pred. No. 2.4e-11;  
Matches 80; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 52 AAAAAACACACAGCAAAACAGTTACAGACCCGTCGAGACTACCTCATCAAACTACTTAGC 111  
AA  
Db 116168 AAAAAACACACAGCAAAACAGTTACAGACCCGTCGAGACTACCTCATCAAACTACTTAGC 116227

QY 112 AGAGATCTTGCAGAGAGAGAGGCTC 136  
AA

Db 116228 AGAGATCTTGCAGAGAGAGAGGCTC 116252

RESULT 10

AC008531

LOCUS

DEFINITION

AC008531

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

COMMENT

AC008531 Homo sapiens chromosome 5 clone CTC-480B11, WORKING DRAFT SEQUENCE,  
7 ordered pieces.  
AC008531  
AC008531.3 GI:12830078  
HTG; HTGS\_PHASE2; HTGS\_DRAFT.  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 145659)  
Sequencing of Human Chromosome 5  
Unpublished  
DOE Joint Genome Institute.  
2 (bases 1 to 145659)  
Direct Submission  
DOE Joint Genome Institute.  
Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
On Feb 14, 2001 this sequence version replaced gi:7528342.

-----Genome Center  
Center: Joint Genome Institute  
Center Code: JGI  
Web site: <http://www.jgi.doe.gov>

```
-----
Project Information
Center Project Name: 369535
Center clone name: CIT-HSPC_480B11
-----
Summary Statistics
Consensus quality: 139128 bases at least Q40
Consensus quality: 142556 bases at least Q30
Consensus quality: 143744 bases at least Q20
Estimated insert size: 148000; pulse field gel estimation
Quality coverage: 6.27 in Q20 bases; sum-of-contigs estimation
Quality coverage: 6.4 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
1
56174: contig of 56174 bp in length
56175 56274: gap of unknown length
56275 100874: contig of 44600 bp in length
100875 100974: gap of unknown length
100975 113127: contig of 12153 bp in length
113128 113227: gap of unknown length
113228 118190: contig of 4963 bp in length
118191 118290: gap of unknown length
118291 119694: contig of 1404 bp in length
119695 119794: gap of unknown length
119795 123297: contig of 3503 bp in length
123298 123398: gap of unknown length
123399 145659: contig of 22262 bp in length.
-----
FEATURES
source
1. .145659
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CITC-480B11"
/clone_lib="Caltech human BAC library C"
BASE COUNT 42561 a 26309 c 27580 g 48609 t 600 others
ORIGIN

Query Match 50.3%; Score 77; DB 2; Length 145659;
Best Local Similarity 94.1%; Pred. No. 2.4e-11;
Matches 80; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 52 AAAAAACCAAGCAAGTACAGCCGTCAGACCTCATCAAACTACTTAGC 111
|||||
Db 46791 AAAAAACCAAGCAAGTACAGCCGTCAGACCTCATCAAACTACTTAGT 46850
|||||

Qy 112 AGAGATCTTGC AAAAGAGGCTC 136
|||||
Db 46851 AGAGATCTTGC AAAAGAGGCTC 46875
|||||

RESULT 11
AC091946 193446 bp DNA linear HTG 09-JUN-2001
DEFINITION Homo sapiens chromosome 5 clone RP11-36012, *** SEQUENCING IN
PROGRESS ***, 33 unordered pieces.
ACCESSION AC091946
VERSION AC091946.1 GI:14333882
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 193446)
AUTHORS DOE Joint Genome Institute.
Sequencing of Human Chromosome 5
Unpublished
2 (bases 1 to 193446)
DOE Joint Genome Institute.
Direct Submission
Submitted (09-JUN-2001) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 544799
Center clone name: RPCI-11_36012
-----
Summary Statistics
Consensus quality: 157767 bases at least Q40
Consensus quality: 180259 bases at least Q30
Consensus quality: 184175 bases at least Q20
Estimated insert size: 204590; agarose-fp estimation
Quality coverage: 5.38 in Q20 bases; agarose-fp estimation
Quality coverage: 5.79 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 33 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
1358: contig of 1358 bp in length
1458: gap of unknown length
1459 2667: contig of 1209 bp in length
2668 2767: gap of unknown length
2768 3823: contig of 1056 bp in length
3824 3923: gap of unknown length
3924 5920: contig of 1997 bp in length
5921 6021: gap of unknown length
6021 7494: contig of 1474 bp in length
7495 7594: gap of unknown length
7595 9096: contig of 1502 bp in length
9097 9196: gap of unknown length
9197 10260: contig of 1064 bp in length
10261 10360: gap of unknown length
10361 12460: contig of 2100 bp in length
12461 12560: gap of unknown length
12561 14611: contig of 2051 bp in length
14612 14711: gap of unknown length
14712 16381: contig of 1670 bp in length
16382 16481: gap of unknown length
16482 17968: contig of 1487 bp in length
17969 18068: gap of unknown length
18069 20434: contig of 2366 bp in length
20435 20534: gap of unknown length
20535 23515: contig of 2981 bp in length
23516 23616: gap of unknown length
23617 27563: contig of 3948 bp in length
27564 27663: gap of unknown length
27664 30987: contig of 3324 bp in length
30988 31087: gap of unknown length
31088 36065: contig of 4978 bp in length
36066 36165: gap of unknown length
36166 40978: contig of 4813 bp in length
40979 41078: gap of unknown length
41079 45663: contig of 4585 bp in length
45664 45763: gap of unknown length
45764 51745: contig of 5982 bp in length
51746 51845: gap of unknown length
51846 57359: contig of 5514 bp in length
57360 57459: gap of unknown length
57460 67881: contig of 10422 bp in length
```

```

* 67882 67981: gap of unknown length
* 67982 74132: contig of 6151 bp in length
* 74133 74232: gap of unknown length
* 74233 79795: contig of 5563 bp in length
* 79796 79895: gap of unknown length
* 79896 87511: contig of 7616 bp in length
* 87512 87611: gap of unknown length
* 87612 92791: contig of 5180 bp in length
* 92792 92891: gap of unknown length
* 92892 102794: contig of 9903 bp in length
* 102795 102894: gap of unknown length
* 102895 110866: contig of 7972 bp in length
* 110867 110966: gap of unknown length
* 110967 117571: contig of 6605 bp in length
* 117572 123738: contig of 6067 bp in length
* 123739 123838: gap of unknown length
* 123839 130583: contig of 6745 bp in length
* 130584 130683: gap of unknown length
* 130684 141544: contig of 10861 bp in length
* 141545 141644: gap of unknown length
* 141645 169109: contig of 27465 bp in length
* 169110 169208: gap of unknown length
* 169210 193446: contig of 24237 bp in length.
FEATURES             Location/Qualifiers
     source            1..193446
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /chromosome="5"
                        /clone_lib="RPCI human BAC library 11"
BASE COUNT           57571 a 35252 c 37061 g 60360 t 3202 others
ORIGIN

Query Match          50.3%; Score 77; DB 2; Length 193446;
Best Local Similarity 94.1%; Pred. No. 2.4e-11;
Matches 80; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 52 AAAAAACCAAGCAAAAGTACAGACCGTGCAGACTACCTCATCAACTACTAGC 111
      |||||||||||||||||||
Db 87679 AAAAAACCAAGCAAAAGTACAGACCGTGCAGACTACCTCATCAACTACTAGC 87738

QY 112 AGAGATCTGCAAAAGAGAGGCTC 136
      |||||||||||||||||||
Db 87739 AGAGATCTGCAAAAGAGAGGCTC 87763

RESULT 12
AC022121/c
LOCUS               219258 bp DNA linear PRI 30-AUG-2001
DEFINITION          Homo sapiens chromosome 5 clone CTD-2007H13, complete sequence.
ACCESSION            AC022121
VERSION              AC022121.6 GI:15375145
KEYWORDS             HTG.
SOURCE               human.
ORGANISM             Homo sapiens
                    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                    Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
  1 (bases 1 to 219258)
  Authors: DOE Joint Genome Institute and Stanford Human Genome Center.
  Title: Direct Submission
  Journal: Unpublished
  2 (bases 1 to 219258)
  Authors: DOE Joint Genome Institute.
  Title: Direct Submission
  Journal: Submitted (26-JAN-2000) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
  3 (bases 1 to 219258)
  Authors: DOE Joint Genome Institute and Stanford Human Genome Center.
  Title: Direct Submission
  Journal: Submitted (11-AUG-2001) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

```

```

REFERENCE
AUTHORS             DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE               Direct Submission
JOURNAL             Submitted (30-AUG-2001) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT             On Aug 30, 2001 this sequence version replaced gi:15148108.
                   Draft Sequence Produced by DOE Joint Genome Institute
                   www.jgi.doe.gov
                   Finishing Completed at Stanford Human Genome Center
                   www.shgc.stanford.edu
                   Quality: Phrap Quality >=40 99.8% of Sequence;
                   Estimated Total Number of Errors is 0.4.
                   STS Content:
                   WI-5811 G04974
                   WI-13675 G23101
                   SHGC-58345 G38487
                   SHGC-103595 G57841
FEATURES             Location/Qualifiers
     source            1..219258
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /chromosome="5"
                        /clones="CTD-2007H13"
BASE COUNT           71954 a 42062 c 40933 g 64309 t
ORIGIN

Query Match          50.3%; Score 77; DB 9; Length 219258;
Best Local Similarity 94.1%; Pred. No. 2.4e-11;
Matches 80; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 52 AAAAAACCAAGCAAAAGTACAGACCGTGCAGACTACCTCATCAACTACTAGC 111
      |||||||||||||||||||
Db 89429 AAAAAACCAAGCAAAAGTACAGACCGTGCAGACTACCTCATCAACTACTAGC 89370

QY 112 AGAGATCTGCAAAAGAGAGGCTC 136
      |||||||||||||||||||
Db 89369 AGAGATCTGCAAAAGAGAGGCTC 89345

RESULT 13
AC092382/c
LOCUS               276181 bp DNA linear HTG 03-JUL-2001
DEFINITION          Homo sapiens chromosome 5 clone RP11-75H1, WORKING DRAFT SEQUENCE,
ACCESSION            AC092382
VERSION              AC092382.1 GI:14589571
KEYWORDS             HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE               human.
ORGANISM             Homo sapiens
                    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                    Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
  1 (bases 1 to 276181)
  Authors: DOE Joint Genome Institute.
  Title: Sequencing of Human Chromosome 5
  Journal: Unpublished
  2 (bases 1 to 276181)
  Authors: DOE Joint Genome Institute.
  Title: Direct Submission
  Journal: Submitted (03-JUL-2001) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
  Center: Joint Genome Institute
  Center Code: JGI
  Web site: http://www.jgi.doe.gov
  Project Information
  Center Project Name: 435334
  Center clone name: RPCI-11_75H1
  Summary Statistics
  Consensus quality: 229677 bases at least Q40
  Consensus quality: 256163 bases at least Q30

```

Consensus quality: 260799 bases at least Q20  
Estimated insert size: 174820; agarose-fp estimation  
Estimated insert size: 271581; sum-of-contigs estimation  
Quality coverage: 8.62 in Q20 bases; agarose-fp estimation  
Quality coverage: 5.55 in Q20 bases; sum-of-contigs estimation.  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 47 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 1061: contig of 1061 bp in length  
\* 1062 1161: contig of unknown length  
\* 1162 2827: contig of 1866 bp in length  
\* 2828 2927: gap of unknown length  
\* 2928 4227: contig of 1300 bp in length  
\* 4228 4327: gap of unknown length  
\* 4328 5618: contig of 1291 bp in length  
\* 5619 5718: gap of unknown length  
\* 5719 5983: contig of 1265 bp in length  
\* 5984 7083: gap of unknown length  
\* 7084 8422: contig of 1339 bp in length  
\* 8423 8522: gap of unknown length  
\* 8523 9771: contig of 1249 bp in length  
\* 9772 9871: gap of unknown length  
\* 9872 11198: contig of 1327 bp in length  
\* 11199 12198: gap of unknown length  
\* 12200 12414: contig of 1116 bp in length  
\* 12415 12514: gap of unknown length  
\* 12515 13806: contig of 1292 bp in length  
\* 13807 13906: gap of unknown length  
\* 13907 15360: contig of 1454 bp in length  
\* 15361 15460: gap of unknown length  
\* 15461 16667: contig of 1207 bp in length  
\* 16668 16767: gap of unknown length  
\* 16768 18390: contig of 1623 bp in length  
\* 18391 18490: gap of unknown length  
\* 18491 19541: contig of 1051 bp in length  
\* 19542 19641: gap of unknown length  
\* 19642 21156: contig of 1515 bp in length  
\* 21157 21256: gap of unknown length  
\* 21257 22448: contig of 1192 bp in length  
\* 22449 22548: gap of unknown length  
\* 22549 24499: contig of 1951 bp in length  
\* 24500 24599: gap of unknown length  
\* 24600 26384: contig of 1785 bp in length  
\* 26385 26484: gap of unknown length  
\* 26485 27883: contig of 1399 bp in length  
\* 27884 27983: gap of unknown length  
\* 27984 30386: contig of 2403 bp in length  
\* 30387 30486: gap of unknown length  
\* 30487 32480: contig of 1994 bp in length  
\* 32481 32580: gap of unknown length  
\* 32581 35441: contig of 2861 bp in length  
\* 35442 35541: gap of unknown length  
\* 35542 37800: contig of 2259 bp in length  
\* 37801 37900: gap of unknown length  
\* 37901 40587: contig of 2887 bp in length  
\* 40588 40687: gap of unknown length  
\* 40688 44159: contig of 3472 bp in length  
\* 44160 44259: gap of unknown length  
\* 44260 46636: contig of 2377 bp in length  
\* 46637 46736: gap of unknown length  
\* 46737 50082: contig of 3346 bp in length  
\* 50083 50182: gap of unknown length  
\* 50183 53988: contig of 3806 bp in length  
\* 53989 54088: gap of unknown length  
\* 54089 56592: contig of 2504 bp in length  
\* 56593 61352: gap of unknown length  
\* 61352 61452: gap of unknown length  
\* 61453

64254: contig of 2802 bp in length  
\* 64255 64354: gap of unknown length  
\* 64355 68825: contig of 4471 bp in length  
\* 68826 68925: gap of unknown length  
\* 68926 74393: contig of 5468 bp in length  
\* 74394 74493: gap of unknown length  
\* 74494 80561: contig of 6068 bp in length  
\* 80562 80661: gap of unknown length  
\* 80662 87626: contig of 6965 bp in length  
\* 87627 87726: gap of unknown length  
\* 87727 93600: contig of 5874 bp in length  
\* 93601 93700: gap of unknown length  
\* 93701 102024: contig of 8324 bp in length  
\* 102025 102124: gap of unknown length  
\* 102125 109005: contig of 7781 bp in length  
\* 109006 110005: gap of unknown length  
\* 110006 116737: contig of 6732 bp in length  
\* 116738 116837: gap of unknown length  
\* 116838 126797: contig of 9960 bp in length  
\* 126798 126897: gap of unknown length  
\* 126898 135114: contig of 8217 bp in length  
\* 135115 135214: gap of unknown length  
\* 135215 146713: contig of 11499 bp in length  
\* 146714 146813: gap of unknown length  
\* 146814 157894: contig of 11081 bp in length  
\* 157895 157994: gap of unknown length  
\* 157995 171752: contig of 13758 bp in length  
\* 171753 171852: gap of unknown length  
\* 171853 201687: contig of 29835 bp in length  
\* 201688 201787: gap of unknown length  
\* 201788 231268: contig of 29481 bp in length  
\* 231269 231368: gap of unknown length  
\* 231369 276181: contig of 44813 bp in length.

FEATURES  
source

Location/Qualifiers  
1. .276181  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="5"  
/clone="RP11-75H1"  
/clone\_lib="RPCI human BAC library 11"  
BASE COUNT 86859 a 51769 c 50026 g 82854 t 4673 others  
ORIGIN

Query Match 50.3%; Score 77; DB 2: Length 276181;  
Best Local Similarity 94.1%; Pred. No. 2.4e-11;  
Matches 80; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 52 AAAAAACCAAGCAAAACAGTTACAGACCCGTCGAGACTACCTCATCAAACTACTAGC 111  
|||||  
Db 213673 AAAAAACCAAGCAAAACAGTTACAGACCCGTCGAGACTACCTCATCAAACTACTAGC 213614

Qy 112 AGAGATCTTGCAAAAAGAGAGCTC 136  
|||||

Db 213613 AGAGATCTTGCAAAAAGAGAGCTC 213589

RESULT 14  
A58691 LOCUS A58691 6608 bp DNA linear PAT 06-MAR-1998  
DEFINITION Sequence 10 from Patent WO9639505.  
ACCESSION A58691  
VERSION A58691.1 GI:3714250  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified  
REFERENCE 1 (bases 1 to 6608)  
AUTHORS Griffiths, R. and Tiwari, B.  
TITLE AVIAN GHD GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN  
BIRDS  
JOURNAL Patent: WO 9639505-A 10 12-DEC-1996;  
ISIS INNOVATION (GB)





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 2, 2002, 20:44:37 ; Search time 1205.34 seconds  
(without alignments)  
217.937 Million cell updates/sec

Title: US-08-973-363-2  
Perfect score: 153  
Sequence: 1 ATTCTTCAGATGATCTGA.....CTCAGAGACTTTGTGTGGCG 153

Scoring table: IDENTITY\_NUC  
Gap 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_032802.\*  
1: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	153	100.0	153	18	AA42756
2	107.2	70.1	153	18	AA42758
3	105.6	69.0	153	18	AA42759
4	102.4	66.9	153	18	AA42757
5	68.4	44.7	6008	18	AA42751
6	67.4	44.1	1316	18	AA42754
7	54	35.3	1311	15	AAV59280
8	41.2	26.9	421	22	AAK88882
9	41.2	26.9	421	22	AA157603

c 10	37.4	24.4	6118	24	ABL33031	Human immune syste
c 11	35.4	23.1	5649	22	AAS46384	Tumour suppressor
c 12	35.4	23.1	5649	24	ABL32849	Human immune syste
c 13	35	22.9	413	22	AA185871	Human polynucleoti
c 14	35	22.9	6283	24	ABL32834	Human immune syste
c 15	34.6	22.6	2944528	24	ABA03041	Listeria monocytog
c 16	34.4	22.5	15644	22	AAS34611	Human DNA for a no
c 17	34.2	22.4	591	20	AA20389	Borrelia burgdorfe
c 18	34.2	22.4	847	20	AA20389	B. burgdorferi ant
c 19	34.2	22.4	847	20	AA20389	B. burgdorferi ant
c 20	34.2	22.4	8318	20	AA20389	Borrelia burgdorfe
c 21	34.2	22.4	11244	22	AA20389	Human immune/hae
c 22	34.2	22.4	11327	22	AA20389	Tumour suppressor
c 23	34.2	22.4	15296	22	AA20389	Human immune/hae
c 24	34.2	22.4	15296	22	AA20389	Human polynucleoti
c 25	33.4	21.8	7255	24	ABL33279	Human immune syste
c 26	33.4	21.8	8513	22	AAS45354	Chemically pretrea
c 27	33.4	21.8	8513	22	AAS45354	Tumour suppressor
c 28	33.2	21.7	238	21	AA52656	Eosinophil activat
c 29	33.2	21.7	939	20	AAZ08845	P-selectin ligand
c 30	33.2	21.7	1239	15	AAQ63954	P-selectin ligand
c 31	33.2	21.7	1239	16	AAQ63954	P-selectin ligand
c 32	33.2	21.7	1239	19	AAV67131	Human glycoprotein
c 33	33.2	21.7	1239	19	AAV67131	CDNA encoding a P-
c 34	33.2	21.7	1239	20	AAZ08860	Human P-selectin l
c 35	33.2	21.7	1239	20	AAV64998	Human P-selectin l
c 36	33.2	21.7	1239	20	AAV69836	Human placenta P-s
c 37	33.2	21.7	1278	20	AAZ08847	P-selectin ligand
c 38	33.2	21.7	1602	20	AAZ08846	P-selectin ligand
c 39	33.2	21.7	1941	20	AAZ08848	P-selectin ligand
c 40	33.2	21.7	3409	21	AAF21427	Human low adenosin
c 41	33.2	21.7	3409	21	AAF21428	Human low adenosin
c 42	33.2	21.7	3409	21	AAA35305	Human adenosine re
c 43	33.2	21.7	3409	21	AAA35306	Human adenosine re
c 44	33.2	21.7	7353	24	ABL32073	Human immune syste
c 45	33.2	21.7	9170	21	AAF21429	Human low adenosin

ALIGNMENTS

RESULT 1  
AAT42756  
ID AAT42756 standard; DNA; 153 BP.

AC AAT42756;  
XX  
XX 12-MAR-1997 (first entry)  
XX  
XX Mouse CHD-1 gene (bases 3855-977).  
XX  
XX Bird; sex determination; chromodomain-Helicase-DNA binding 1;  
XX  
XX CHD-1; CHD-W; w chromosome; ss.

OS Mus sp.

XX  
XX Key Location/Qualifiers

FT misc\_difference 52..81

FT /\*tag= a

FT /note= "bases 52-81 are a repeat of bases 22-51

FT and are ignored in the translated amino

FT acid sequence given in Fig 3"

XX  
XX W09639505-A1.

XX  
XX 12-DEC-1996.

XX  
XX 05-JUN-1996; 96WO-GB01341.

XX  
XX 06-JUN-1995; 95GB-0011439.

XX  
XX (ISIS-) ISIS INNOVATION LTD.

```
PI Griffiths R, Tiwari B;
XX WPI; 1997-043127/04.
DR P-PSDB; AAW08146.
XX Avian chromodomain-helicase-DNA binding genes determine sex in
PT birds - used for sex determin. and to control sex of progeny
XX Claim 8; Fig 3; 76pp; English.
PS Bases 3855-3977 (AAT42756) of the mouse CHD-1 gene show homology
CC to portions of the chicken CHD-1A (A = Avian) gene (AAT42757),
CC and AAT42757), chicken CHD-W (W refers to the W chromosome) gene
CC (see also AAT42758) and the great tit CHD-W gene (see also AAT42759).
CC translated amino acid sequences of this region are provided in
CC AAW08146-49. The CHD-1A (see also AAT42751) and CHD-W (see also
CC AAT42754-55) genes determine sex in birds and can be used to identify
CC the sex of an embryo, foetus etc. and to manipulate the sex of
XX progeny.
XX Sequence 153 BP; 61 A; 37 C; 27 G; 28 T; 0 other;
SQ
Query Match 100.0%; Score 153; DB 18; Length 153;
Best Local Similarity 100.0%; Pred. No. 1.8e-34;
Matches 153; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATTCTTCCAGATGATCCTCATATAAAACACCAAGCAAAACAGTTACAGACCAAAAACCA 60
DB 1 attcttcagatgactcctgataaaacacacagcaaacagttacagaccaaataa 60
QY 61 CAAGCAAAACAGTTACAGACCGTGCAGACTACCTCATCAAACTACTTAGCAGAGATCTT 120
DB 61 caagcaaaacagttacagacccgtgcagactacctcatcaactacttagcagagatctt 120
QY 121 GCAAAAAGAGAGCGCTCAGAGACTTTGTGGTGG 153
DB 121 gcaaaaagagagctcagagactttgtggtg 153
RESULT 2
AAT42759
ID AAT42759 standard; DNA; 153 BP.
XX AC AAT42759;
XX DT 12-MAR-1997 (first entry)
XX DE Great tit CHD-W gene fragment.
XX KW Bird; sex determination; chromodomain-Helicase-DNA binding 1;
XX KW CHD-1A; CHD-W; W chromosome; ss.
XX OS Parus major.
XX FH Key Location/Qualifiers
XX FT misc_difference 52..81
XX FT /*tag= a
XX FT /note= "bases 52-81 are a repeat of bases 22-51
XX FT and are ignored in the translated amino
XX FT acid sequence given in Fig 3"
XX PN WO9639505-A1.
XX PD 12-DEC-1996.
XX PF 05-JUN-1996; 96WO-GB01341.
XX PR 06-JUN-1995; 95GB-0011439.
XX PA (ISIS-) ISIS INNOVATION LTD.
XX PI Griffiths R, Tiwari B;
DR WPI; 1997-043127/04.
```

```
XX WPI; 1997-043127/04.
DR P-PSDB; AAW08149.
XX Avian chromodomain-helicase-DNA binding genes determine sex in
PT birds - used for sex determin. and to control sex of progeny
XX Claim 8; Fig 3; 76pp; English.
XX Bases 3855-3977 (AAT42756) of the mouse CHD-1 gene show homology
CC to portions of the chicken CHD-1A (A = Avian) gene (AAT42757),
CC chicken CHD-W (W refers to the W chromosome) gene (AAT42758) and
CC and the great tit CHD-W gene (AAT42759). Translated amino acid
CC sequences of this region are provided in AAW08146-49. The CHD-1A
CC (see also AAT42751) and CHD-W (see also AAT42754-55) genes determine
CC sex in birds and can be used to identify the sex of an embryo,
CC foetus etc. and to manipulate the sex of progeny.
XX Sequence 153 BP; 60 A; 33 C; 31 G; 29 T; 0 other;
SQ
Query Match 70.1%; Score 107.2; DB 18; Length 153;
Best Local Similarity 81.6%; Pred. No. 2e-21;
Matches 124; Conservative 0; Mismatches 28; Indels 0; Gaps 0;
QY 1 ATTCTTCCAGATGATCCTCATATAAAACACCAAGCAAAACAGTTACAGACCAAAAACCA 60
DB 1 atttcttcagatgactcctgataaaacacacagcaaacagttgcagaccaaagaacca 60
QY 61 CAAGCAAAACAGTTACAGACCGTGCAGACTACCTCATCAAACTACTTAGCAGAGATCTT 120
DB 61 caggcaaaagcagttgcagacccgtgcagactaccttaataattactgaataaagacctt 120
QY 121 GCAAAAAGAGAGCGCTCAGAGACTTTGTGGTGC 152
DB 121 gcaaaaagagagctcagagactttgtggtgc 152
RESULT 3
AAT42758
ID AAT42758 standard; DNA; 153 BP.
XX AC AAT42758;
XX DT 12-MAR-1997 (first entry)
XX DE Chick CHD-W gene fragment.
XX KW Bird; sex determination; chromodomain-Helicase-DNA binding 1;
XX KW CHD-1A; CHD-W; W chromosome; ss.
XX OS Gallus sp.
XX FH Key Location/Qualifiers
XX FT misc_difference 52..81
XX FT /*tag= a
XX FT /note= "bases 52-81 are a repeat of bases 22-51
XX FT and are ignored in the translated amino
XX FT acid sequence given in Fig 3"
XX PN WO9639505-A1.
XX PD 12-DEC-1996.
XX PF 05-JUN-1996; 96WO-GB01341.
XX PR 06-JUN-1995; 95GB-0011439.
XX PA (ISIS-) ISIS INNOVATION LTD.
XX PI Griffiths R, Tiwari B;
XX WPI; 1997-043127/04.
```



DR P-PSDB; AAW08148.  
 XX Avian chromodomain-helicase-DNA binding genes determine sex in  
 PT birds - used for sex determin. and to control sex of progeny  
 XX  
 PS Claim 8; Fig 3; 76pp; English.  
 XX  
 CC Bases 3855-3977 (AAT42756) of the mouse CHD-1 gene show homology  
 CC to portions of the chicken CHD-1A (A = Avian) gene (AAT42757),  
 CC chicken CHD-W (W refers to the W chromosome) gene (AAT42758) and  
 CC the great tit CHD-W gene (see also AAT42759). Translated amino acid  
 CC sequences of this region are provided in AAW08146-49. The CHD-1A  
 CC (see also AAT42751) and CHD-W (see also AAT42754-55) genes determine  
 CC sex in birds and can be used to identify the sex of an embryo,  
 CC foetus etc. and to manipulate the sex of progeny.  
 XX  
 SQ Sequence 153 BP; 56 A; 36 C; 31 G; 30 T; 0 other;

Query Match 69.0%; Score 105.6; DB 18; Length 153;  
 Best Local Similarity 80.9%; Pred. No. 5.8e-21;  
 Matches 123; Conservative 0; Mismatches 29; Indels 0; Gaps 0;  
 Qy 1 ATTCTCCAGATGATCCTGATAAAACCAACCAAGCAAAACAGTTTACAGACCAAAACCA 60  
 Db 1 atttaccctgatgcagataaagaacccagcgttaagcagttacagaccagaacccc 60  
 Qy 61 CAAGCAAAACAGTTACAGACCCGTCAGACTACCTCATCAAACTACTTACAGAGATCTT 120  
 Db 61 caggcttaagcagttacagaccgctgcagattaccttaataattactgaataaagacctt 120  
 Qy 121 GCAAAAAGAGAGGCTCAGAGACTTTGTGGTGC 152  
 Db 121 gcaagaagaagcacagagacttgcgtgtgc 152

RESULT 4  
 AAT42757  
 ID AAT42757 standard; DNA; 153 BP.  
 XX  
 AC AAT42757;  
 XX  
 DT 12-MAR-1997 (first entry)  
 XX  
 DE Chick CHD-1A gene fragment.  
 XX  
 KW Bird; sex determination; chromodomain-Helicase-DNA binding 1;  
 KW CHD-1A; CHD-W; W chromosome; ss.  
 XX  
 OS Gallus sp.  
 XX  
 FH Key Location/Qualifiers  
 FT misc\_difference 52..81  
 FT /\*tag= a  
 FT /note= "bases 52-81 are a repeat of bases 22-51  
 FT and are ignored in the translated amino  
 FT acid sequence given in Fig 3"  
 XX  
 PN WO9639505-A1.  
 XX  
 PD 12-DEC-1996.  
 XX  
 PF 05-JUN-1996; 96WO-GB01341.  
 XX  
 PR 06-JUN-1995; 95GB-0011439.  
 XX  
 PA (ISIS-) ISIS INNOVATION LTD.  
 XX  
 PI Griffiths R, Tiwari B;  
 XX  
 DR WPI; 1997-043127/04.  
 DR P-PSDB; AAW08147.  
 XX

PT Avian chromodomain-helicase-DNA binding genes determine sex in  
 XX birds - used for sex determin. and to control sex of progeny  
 XX  
 PS Claim 8; Fig 3; 76pp; English.  
 XX  
 CC Bases 3855-3977 (AAT42756) of the mouse CHD-1 gene show homology  
 CC to portions of the chicken CHD-1A (A = Avian) gene (AAT42757),  
 CC chicken CHD-W (W refers to the W chromosome) gene (see also AAT42758)  
 CC and the great tit CHD-W gene (see also AAT42759). Translated amino  
 CC acid sequences of this region are provided in AAW08146-49. The  
 CC CHD-1A (see also AAT42751) and CHD-W (see also AAT42754-55) genes  
 CC determine sex in birds and can be used to identify the sex of an  
 CC embryo, foetus etc. and to manipulate the sex of progeny.  
 XX  
 SQ Sequence 153 BP; 58 A; 40 C; 31 G; 24 T; 0 other;

Query Match 56.9%; Score 102.4; DB 18; Length 153;  
 Best Local Similarity 79.6%; Pred. No. 4.7e-20;  
 Matches 121; Conservative 0; Mismatches 31; Indels 0; Gaps 0;  
 Qy 1 ATTCTCCAGATGATCCTGATAAAACCAACCAAGCAAAACAGTTTACAGACCAAAACCA 60  
 Db 1 atttaccctgatgcagataaagaacccagcgttaagcagttacagaccagaacccc 60  
 Qy 61 CAAGCAAAACAGTTACAGACCCGTCAGACTACCTCATCAAACTACTTACAGAGATCTT 120  
 Db 61 caggcaagcagctacagaccgctgcagactacctcattaaattactgaataaagacctt 120  
 Qy 121 GCAAAAAGAGAGGCTCAGAGACTTTGTGGTGC 152  
 Db 121 gcaagaagaagcacacaaagcgttgcgtgtgc 152

RESULT 5  
 AAT42751  
 ID AAT42751 standard; CDNA; 6608 BP.  
 XX  
 AC AAT42751;  
 XX  
 DT 12-MAR-1997 (first entry)  
 XX  
 DE Chicken CHD-1A gene.  
 XX  
 KW Bird; sex determination; chromodomain-Helicase-DNA binding 1 Avian;  
 KW CHD-W; chromodomain-Helicase-DNA binding on the W chromosome; ss.  
 XX  
 OS Gallus sp.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 228..5390  
 FT /\*tag= a  
 XX  
 PN WO9639505-A1.  
 XX  
 PD 12-DEC-1996.  
 XX  
 PF 05-JUN-1996; 96WO-GB01341.  
 XX  
 PR 06-JUN-1995; 95GB-0011439.  
 XX  
 PA (ISIS-) ISIS INNOVATION LTD.  
 XX  
 PI Griffiths R, Tiwari B;  
 XX  
 DR WPI; 1997-043127/04.  
 XX  
 PT Avian chromodomain-helicase-DNA binding genes determine sex in  
 PT birds - used for sex determin. and to control sex of progeny  
 XX  
 PS Claim 1; Fig 5; 76pp; English.  
 XX  
 CC The chicken CHD-W gene (AAT42754) acting alone or in conjunction with



```

Query Match      35.3%; Score 54; DB 19; Length 1311;
Best Local Similarity 100.0%; Pred. No. 4.9e-06;
Matches 54; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 52 AAAAAACCAAGCAAAACAGTTACAGACCCGTCGAGACTACCTCATCAACTA 105
    |||||
Db 1258 aaaaacacaaagcaaaacagttacagaccgctgcagactacatcaacta 1311

RESULT      8
AAK88882
ID AAK88882 standard; cDNA; 421 BP.
XX
AC AAK88882;
XX
DX 05-NOV-2001 (first entry)
DE Human digestive system antigen coding sequence SEQ ID NO: 1198.
DE
DE Human; digestive system antigen; gene therapy; cancer; appendicitis;
KW ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;
KW digestive system disorder; Meckel's diverticulum; ss.
XX
OS Homo sapiens.
XX
XX WO200155314-A2.
XX
XX 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01324.
XX
XX 31-JAN-2000; 2000US-0179065.
XX 04-FEB-2000; 2000US-0180628.
XX 24-FEB-2000; 2000US-0184664.
XX 02-MAR-2000; 2000US-0186350.
XX 16-MAR-2000; 2000US-0189874.
XX 17-MAR-2000; 2000US-0190076.
XX 18-APR-2000; 2000US-0198123.
XX 19-MAY-2000; 2000US-0205515.
XX 07-JUN-2000; 2000US-0209467.
XX 28-JUN-2000; 2000US-0214886.
XX 30-JUN-2000; 2000US-0215139.
XX 07-JUL-2000; 2000US-0215139.
XX 07-JUL-2000; 2000US-0215647.
XX 11-JUL-2000; 2000US-0215680.
XX 11-JUL-2000; 2000US-02117487.
XX 14-JUL-2000; 2000US-02117496.
XX 14-JUL-2000; 2000US-0218290.
XX 26-JUL-2000; 2000US-0220963.
XX 14-AUG-2000; 2000US-0220964.
XX 14-AUG-2000; 2000US-0224518.
XX 14-AUG-2000; 2000US-0224519.
XX 14-AUG-2000; 2000US-0225213.
XX 14-AUG-2000; 2000US-0225214.
XX 14-AUG-2000; 2000US-0225266.
XX 14-AUG-2000; 2000US-0225267.
XX 14-AUG-2000; 2000US-0225268.
XX 14-AUG-2000; 2000US-0225270.
XX 14-AUG-2000; 2000US-0225477.
XX 14-AUG-2000; 2000US-0225757.
XX 14-AUG-2000; 2000US-0225758.
XX 14-AUG-2000; 2000US-0225759.
XX 14-AUG-2000; 2000US-0226279.
XX 22-AUG-2000; 2000US-0226681.
XX 22-AUG-2000; 2000US-0226688.
XX 22-AUG-2000; 2000US-0227182.
XX 23-AUG-2000; 2000US-0227009.
XX 30-AUG-2000; 2000US-0228924.
XX 01-SEP-2000; 2000US-0229287.
XX 01-SEP-2000; 2000US-0229343.
XX 01-SEP-2000; 2000US-0229344.
XX 01-SEP-2000; 2000US-0229345.
XX 05-SEP-2000; 2000US-0229509.

```

```

PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234423.
PR 21-SEP-2000; 2000US-0234474.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241121.
PR 20-OCT-2000; 2000US-02411783.
PR 20-OCT-2000; 2000US-02411786.
PR 20-OCT-2000; 2000US-02411787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.

```



PR 27-SEP-2000; 2000US-0235836.  
 PR 29-SEP-2000; 2000US-0236327.  
 PR 29-SEP-2000; 2000US-0236367.  
 PR 29-SEP-2000; 2000US-0236368.  
 PR 29-SEP-2000; 2000US-0236369.  
 PR 29-SEP-2000; 2000US-0236370.  
 PR 02-OCT-2000; 2000US-0236802.  
 PR 02-OCT-2000; 2000US-0237037.  
 PR 02-OCT-2000; 2000US-0237038.  
 PR 02-OCT-2000; 2000US-0237039.  
 PR 02-OCT-2000; 2000US-0237040.  
 PR 13-OCT-2000; 2000US-0239935.  
 PR 13-OCT-2000; 2000US-0239937.  
 PR 20-OCT-2000; 2000US-0240960.  
 PR 20-OCT-2000; 2000US-0241221.  
 PR 20-OCT-2000; 2000US-0241785.  
 PR 20-OCT-2000; 2000US-0241786.  
 PR 20-OCT-2000; 2000US-0241787.  
 PR 20-OCT-2000; 2000US-0241808.  
 PR 20-OCT-2000; 2000US-0241809.  
 PR 20-OCT-2000; 2000US-0241826.  
 PR 01-NOV-2000; 2000US-0244617.  
 PR 08-NOV-2000; 2000US-0246474.  
 PR 08-NOV-2000; 2000US-0246475.  
 PR 08-NOV-2000; 2000US-0246476.  
 PR 08-NOV-2000; 2000US-0246477.  
 PR 08-NOV-2000; 2000US-0246478.  
 PR 08-NOV-2000; 2000US-0246523.  
 PR 08-NOV-2000; 2000US-0246524.  
 PR 08-NOV-2000; 2000US-0246525.  
 PR 08-NOV-2000; 2000US-0246526.  
 PR 08-NOV-2000; 2000US-0246527.  
 PR 08-NOV-2000; 2000US-0246528.  
 PR 08-NOV-2000; 2000US-0246529.  
 PR 08-NOV-2000; 2000US-0246603.  
 PR 08-NOV-2000; 2000US-0246610.  
 PR 08-NOV-2000; 2000US-0246611.  
 PR 17-NOV-2000; 2000US-0246613.  
 PR 17-NOV-2000; 2000US-0249207.  
 PR 17-NOV-2000; 2000US-0249208.  
 PR 17-NOV-2000; 2000US-0249209.  
 PR 17-NOV-2000; 2000US-0249210.  
 PR 17-NOV-2000; 2000US-0249211.  
 PR 17-NOV-2000; 2000US-0249212.  
 PR 17-NOV-2000; 2000US-0249213.  
 PR 17-NOV-2000; 2000US-0249214.  
 PR 17-NOV-2000; 2000US-0249215.  
 PR 17-NOV-2000; 2000US-0249216.  
 PR 17-NOV-2000; 2000US-0249217.  
 PR 17-NOV-2000; 2000US-0249218.  
 PR 17-NOV-2000; 2000US-0249244.  
 PR 17-NOV-2000; 2000US-0249245.  
 PR 17-NOV-2000; 2000US-0249264.  
 PR 17-NOV-2000; 2000US-0249265.  
 PR 17-NOV-2000; 2000US-0249297.  
 PR 17-NOV-2000; 2000US-0249299.  
 PR 17-NOV-2000; 2000US-0249300.  
 PR 01-DEC-2000; 2000US-0250160.  
 PR 01-DEC-2000; 2000US-0250391.  
 PR 05-DEC-2000; 2000US-0251030.  
 PR 05-DEC-2000; 2000US-0251988.  
 PR 05-DEC-2000; 2000US-0256719.  
 PR 06-DEC-2000; 2000US-0251479.  
 PR 08-DEC-2000; 2000US-0251856.  
 PR 08-DEC-2000; 2000US-0251868.  
 PR 08-DEC-2000; 2000US-0251869.  
 PR 08-DEC-2000; 2000US-0251869.  
 PR 08-DEC-2000; 2000US-0251989.  
 PR 11-DEC-2000; 2000US-0251990.  
 PR 05-JAN-2001; 2001US-0259678.  
 (HUMA-) HUMAN GENOME SCI INC.

PI Rosen CA, Barash SC, Ruben SM;  
 XX WPI; 2001-457727/49.  
 DR P-PSDB; AAM38625.  
 XX Isolated polypeptide for treating, preventing and/or prognosing  
 PT disorders related to the colon and rectum including colorectal cancers  
 PT and also for testing and detection e.g. diagnosis  
 XX Claim 1; SEQ ID NO: 67; 522pp + Sequence Listing; English.  
 PS The present invention provides the protein and coding sequences of a  
 CC number of colorectal cancer antigens. These are shown in  
 CC AAI5747-AA157619 and AAM38569-AAM38641. These can be used in the  
 CC diagnosis, prevention and treatment of cancer of the colon and/or rectum.  
 CC The present sequence is a colorectal cancer antigen coding sequence of  
 CC the invention.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX Sequence 421 BP; 148 A; 78 C; 127 G; 61 T; 7 other;  
 SQ  
 Query Match 26.9%; Score 41.2; DB 22; Length 421;  
 Best Local Similarity 62.7%; Pred. No. 0.016;  
 Matches 64; Conservative 0; Mismatches 38; Indels 0; Gaps 0;  
 QY 52 AAAAAACCAAGCAACAGTACAGCCCGTCGACAGCTACCTATCAAACTACTTACG 111  
 ||||| || || || || || || || || || || || || || || || || || || || ||  
 Db 131 aaaaagcctcagggaagcagctacagaccgcagcgattactgttgagctgtcaga 190  
 QY 112 AGAGATCTTCCAAAAGAGAGGCTCAGACACTTCTGTCGCG 153  
 || || || || || || || || || || || || || || || || || || || || || || ||  
 Db 191 aagggctctggagaagaagggggctgtgacaggtgagggaagag 232  
 RESULT 10  
 ABL33031/c  
 ID ABL33031 standard; DNA; 6118 BP.  
 XX ABL33031;  
 AC ABL33031;  
 XX 26-MAR-2002 (first entry)  
 DT Human immune system associated gene SEQ ID NO: 1004.  
 XX Human; immune system disease; cytosine methylation; antiasthmatic;  
 KW antiarteriosclerotic; antihaemic; cytostatic; neotropic;  
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;  
 KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;  
 KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;  
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;  
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;  
 XX gene; ds.  
 XX Homo sapiens.  
 XX WO200200928-A2.  
 PN 03-JAN-2002.  
 PD 02-JUL-2001; 2001WO-EP07537.  
 PF 30-JUN-2000; 2000DE-1032529..  
 PR 01-SEP-2000; 2000DE-1043826.  
 XX (EPIG-) EPIGENOMICS AG.  
 PA Olek A, Piepenbrock C, Berlin K;  
 PI WPI; 2002-130909/17.  
 XX

PT Nucleic acid comprising fragment of chemically modified gene, useful  
PT for diagnosis and treatment of diseases associated with abnormal  
PT cytosine methylation -  
XX  
PS Claim 1; SEQ ID NO 1004; 32pp + Sequence Listing; German.  
XX  
CC The present invention provides a number of human immune system associated  
CC genes which are modified by the methylation of cytosines. The sequences  
CC can be used in the diagnosis and treatment of immune system disorders,  
CC including eye diseases such as retinopathy, neovascular glaucoma and  
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
CC leukemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
CC diseases. The present sequence is a gene of the invention.  
XX  
SQ Sequence 6118 BP; 1751 A; 141 C; 1229 G; 2997 T; 0 other;  
  
Query Match 24.4%; Score 37.4; DB 24; Length 6118;  
Best Local Similarity 55.9%; Pred. No. 0.38;  
Matches 71; Conservative 0; Mismatches 56; Indels 0; Gaps 0;  
  
QY 5 TTCCAGATGATCCTGATATAAAACACACAGCAAAACAGTTACAGACCAAAACACACAG 64  
DB 4251 TTCAATAAAATCTAATAAAACACACAAATAAAATACATAAAACCAAAATCAATA 4192  
  
QY 65 CAAAACAGTTACAGACCCGTGCAGACTACCTCATCAAACTACTTAGCAGAGATCTTGCAA 124  
DB 4191 CGAAAAATATAAAACACAAATATCTATATCTAATACTACTACTATACATATACAAT 4132  
  
QY 125 AAAGAGA 131  
DB 4131 AAAAACA 4125  
  
RESULT 11  
AAS46384/c  
ID AAS46384 standard; DNA; 5649 BP.  
AC AAS46384;  
XX  
XX  
DT 18-DEC-2001 (first entry)  
DE Tumour suppressor gene derived chemically modified sequence #106.  
KW Human; tumour suppressor gene; oncogene; antitumour; cytostatic;  
KW cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;  
KW cytosine methylation; ds.  
XX  
OS Homo sapiens.  
XX  
XX WO200168912-A2.  
XX  
PD 20-SEP-2001.  
XX  
PF 15-MAR-2001; 2001WO-EP02955.  
XX  
XX 15-MAR-2000; 2000DE-1013847.  
PR 06-APR-2000; 2000DE-1019058.  
PR 07-APR-2000; 2000DE-1019173.  
PR 30-JUN-2000; 2000DE-1032529.  
PR 01-SEP-2000; 2000DE-1043826.  
XX  
XX (EPIG-) EPIGENOMICS AG.  
PA  
PI Olek A, Piepenbrock C, Berlin K;  
XX  
XX WPI; 2001-602752/68.  
XX  
XX Fragments of chemically modified genes associated with tumour suppressor  
PT genes and oncogenes, useful in designing primers and probes for  
PT analysing diseases associated with cytosine methylation state e.g.  
PT cancer -

XX  
PS Claim 1; SEQ ID NO 106; 27pp; English.  
XX  
CC The invention relates to a nucleic acid comprising a sequence of 18  
CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with  
CC bisulphite, of genes associated with tumour suppression and  
CC oncogenes having a sequence taken from 536 (actually 533 since  
CC numbers 408, 458 and 500 are missing from the sequence listing) sequences  
CC (SS) and sequences complementary to (SS). The nucleic acid may be a  
CC peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may  
CC form part of a set of probes for detecting the cytosine methylation state  
CC and/or single nucleotide polymorphisms and also to be used in an  
CC array for analysing diseases associated with CpG dinucleotides e.g.  
CC cancers and tumours. The probes can also be used in a method for  
CC ascertaining genetic and/or epigenetic parameters for the diagnosis  
CC and/or therapy of existing diseases or the predisposition to specific  
CC diseases, by analysing cytosine methylations. The parameters may be  
CC compared to another set of genetic and/or epigenetic parameters, the  
CC differences serving as basis for diagnosis and/or prognosis events which  
CC are disadvantageous to patients. The present sequence is one of the  
CC 533 genomic sequences derived from tumour suppressor genes and  
CC oncogenes. Sequences with even numbered Seq ID numbers are the  
CC complementary sequence of the corresponding odd numbered sequence (e.g.  
CC ID 2 and ID1, ID 536 and ID 535, except for those whose partner sequence  
CC is missing).  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 5649 BP; 1212 A; 184 C; 1397 G; 2856 T; 0 other;  
  
Query Match 23.1%; Score 35.4; DB 22; Length 5649;  
Best Local Similarity 55.2%; Pred. No. 1.4;  
Matches 69; Conservative 0; Mismatches 56; Indels 0; Gaps 0;  
  
QY 3 TCTTCCAGATGATCTGTATATAAAACCAACCAAGCAAAACAGTTACAGACCAAAACCAACA 62  
DB 3870 TTTTCCAAATATACAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAA 3811  
  
QY 63 AGCAAAACAGTTACAGACCCGTGCAGACTACCTCATCAAACTACTTAGCAGAGATCTTGC 122  
DB 3810 AAAAAACGAAAAAATAAATCTCAAAAAAATAAATAAATAAATACATTTCTCAAAACTAAA 3751  
  
QY 123 AAAAA 127  
DB 3750 AAAAA 3746  
  
RESULT 12  
ABL32849/c  
ID ABL32849 standard; DNA; 5649 BP.  
XX  
AC ABL32849;  
XX  
XX 26-MAR-2002 (first entry)  
DT  
DE Human immune system associated gene SEQ ID NO: 822.  
XX  
XX Human; immune system disease; cytosine methylation; antiasthmatic;  
KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;  
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;  
KW antirheumatic; antiarthritis; antidiabetic; antipsoriatic;  
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;  
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;  
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;  
KW gene; ds.  
XX  
XX Homo sapiens.  
OS  
XX WO200200928-A2.  
PN  
XX

PD 03-JAN-2002.  
 XX  
 XX 02-JUL-2001; 2001WO-EP07537.  
 XX  
 XX 30-JUN-2000; 2000DE-1032529.  
 PR 01-SEP-2000; 2000DE-1043826.  
 XX  
 XX (EPIC-) EPIGENOMICS AG.  
 XX  
 XX Olek A, Piepenbrock C, Berlin K;  
 XX WPI; 2002-130909/17.  
 XX  
 XX Nucleic acid comprising fragment of chemically modified gene, useful  
 PT for diagnosis and treatment of diseases associated with abnormal  
 PT cytosine methylation -  
 XX  
 PS Claim 1; SEQ ID NO 822; 32pp + Sequence Listing; German.  
 XX  
 CC The present invention provides a number of human immune system associated  
 CC genes which are modified by the methylation of cytosines. The sequences  
 CC can be used in the diagnosis and treatment of immune system disorders,  
 CC including eye diseases such as retinopathy, neovascular glaucoma and  
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
 CC diseases. The present sequence is a gene of the invention.  
 XX  
 SQ Sequence 5649 BP; 1212 A; 184 C; 1397 G; 2856 T; 0 other;

Query Match 23.1%; Score 35.4; DB 24; Length 5649;  
 Best Local Similarity 55.2%; Pred. No. 1.4;  
 Matches 69; Conservative 0; Mismatches 56; Indels 0; Gaps 0;  
 QY 3 TCTTCAGATGATCCTGATATAAAACACAGCAAGCAAGTTCACAGACCAAAACACCA 62  
 DB 3870 TTTTCCAAATATACAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 3811  
 QY 63 AGCAAAACAGTTCAGACCCGTCAGACTACCTCAAACTACTTACGAGAGATCTTGG 122  
 DB 3810 AAAAAACGAAAAAATAAAATCTCAAAAAAATAAAATACATTTCTCAAACTAAA 3751  
 QY 123 AAAAA 127  
 DB 3750 AAAAA 3746

RESULT 13  
 AAI85871/C  
 ID AAI85871 standard; cDNA; 413 BP.  
 XX  
 AC AAI85871;  
 XX  
 DT 06-NOV-2001 (first entry)  
 XX  
 DE Human polynucleotide SEQ ID NO 5931.  
 XX  
 KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
 KW nervous system disorders; arthritis; inflammation; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200164835-A2.  
 XX  
 PD 07-SEP-2001.  
 XX  
 XX 26-FEB-2001; 2001WO-US04927.  
 PF  
 XX 28-FEB-2000; 2000US-0515126.  
 PR 18-MAY-2000; 2000US-0577409.  
 PR

XX (HYSE-) HYSEQ INC.  
 XX  
 XX Tang YT, Liu C, Drmanac RT;  
 XX WPI; 2001-514838/56.  
 DR P-PSDB; AAO05940.  
 XX  
 XX Isolated nucleic acids and polypeptides, useful for preventing  
 PT diagnosing and treating e.g. leukaemia, inflammation and immune  
 PT disorders -  
 XX  
 XX Claim 1; SEQ ID NO 5931; 1399pp + Sequence Listing; English.  
 PS  
 XX The invention relates to human polynucleotides (AAI79941-AAI93841) and  
 CC the encoded proteins (AAO0010-AAO13910) that exhibit activity relating to  
 CC cytokine, cell proliferation or cell differentiation or which may induce  
 CC production of other cytokines in other cell populations. The  
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
 CC peptide therapy. The polypeptides have various cytokine-like activities,  
 CC e.g. stem cell growth factor activity, haematopoiesis regulating  
 CC activity, tissue growth factor activity, immunomodulatory activity and  
 CC activin/inhibin activity and may be useful in the diagnosis and/or  
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
 CC inflammation.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 413 BP; 82 A; 22 C; 74 G; 235 T; 0 other;

Query Match 22.9%; Score 35; DB 22; Length 413;  
 Best Local Similarity 54.2%; Pred. No. 0.94;  
 Matches 71; Conservative 0; Mismatches 60; Indels 0; Gaps 0;  
 QY 1 ATTCTTCCAGATGATCCTGATATAAAACCAAGCAAGCAAGTTCACAGACCAAAACCA 60  
 DB 284 ACTATACAAACAACTCATATAAAACCAACAAACCAACCAACCAACCAACCAACCA 225  
 QY 61 CAAGCAAAACAGTTCAGACCCGTCAGACTACCTCAAACTACTTACGAGAGATCTT 120  
 DB 224 AAAAAACAAAAAACAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAA 165  
 QY 121 GCAAAAGAGA 131  
 DB 164 AAAAAAATAA 154

RESULT 14  
 ABL32834/C  
 ID ABL32834 standard; DNA; 6283 BP.  
 XX  
 AC ABL32834;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Human immune system associated gene SEQ ID NO: 807.  
 XX  
 KW Human; immune system disease; cytosine methylation; antiasthmatic;  
 KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;  
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;  
 KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;  
 KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;  
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;  
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;  
 KW gene; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200200928-A2.  
 XX  
 PD 03-JAN-2002.

```
XX 02-JUL-2001; 2001WO-EP07537.
XX 30-JUN-2000; 2000DE-1032529.
XX 01-SEP-2000; 2000DE-1043826.
XX (EPIG-) EPIGENOMICS AG.
XX Olek A, Piepenbrock C, Berlin K;
XX WPI; 2002-130909/17.
XX Nucleic acid comprising fragment of chemically modified gene, useful
XX for diagnosis and treatment of diseases associated with abnormal
XX cytosine methylation -
XX Claim 1; SEQ ID NO 807; 32pp + Sequence Listing; German.
XX The present invention provides a number of human immune system associated
XX genes which are modified by the methylation of cytosines. The sequences
XX can be used in the diagnosis and treatment of immune system disorders,
XX including eye diseases such as retinopathy, neovascular glaucoma and
XX macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
XX leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
XX rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
XX diseases. The present sequence is a gene of the invention.
XX Sequence 6283 BP; 1407 A; 230 C; 1615 G; 3031 T; 0 other;

Query Match      22.9%; Score 35; DB 24; Length 6283;
Best Local Similarity 55.3%; Pred. NO. 1.8;
Matches 68; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 4 CTTCAGATGCTCGTATATAAACCACAGCAACAGTTCACAGACCAAAACCAACCA 63
Db 138 CTTTACATATATATATATCAAAAAAATAAATAACCAATTAATAATTTTACCCAACT 79

QY 64 GCAAAACAGTTACAGACCGCTGCAGACTACCTCATCAAACTACTTAGCAGAGATCTTGA 123
Db 78 ATAAACCTTACCGAACCAGTACCTTCTCTTAACAACTAAACGAATTTTAA 19

QY 124 AAA 126
Db 18 AAA 16

RESULT 15
ABA03041/c
ID ABA03041 standard; DNA; 2944528 BP.
XX ABA03041;
XX 05-FEB-2002 (first entry)
XX Listeria monocytogenes EGD-e genome sequence.
XX Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;
XX vitamin B12; bacterial infection; disease; ds.
XX Listeria monocytogenes.
XX WO200177335-A2.
XX 18-OCT-2001.
XX 11-APR-2001; 2001WO-FR01118.
XX 11-APR-2000; 2000FR-0004629.
XX (INSP ) INST PASTEUR.
XX Buchrieser C, Frangeul L, Couve E, Rusniok C, Fsihi H, Dehoux P;
```

```
PI Dussurget O, Chetouani F, Nedjari H, Glaser P, Kunst F, Cossart P;
PI Daniels J, Goebel W, Krefit J, Kuhn M, Ng E, Vazquez-Boland JA;
PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;
PI Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L;
PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;
PI Maduenio E, De Pablo B, Wehland J, Kaerst U, Entian K, Hauf J;
PI Rose M, Voss H;
XX WPI; 2002-010914/01.
XX Genomic sequence for Listeria monocytogenes, useful e.g. for treatment
XX and prevention of Listeria and related bacterial infections, and
XX related polypeptides -
XX Claim 1; SEQ ID NO 1; 192pp; French.
XX The present sequence is the genome sequence of Listeria monocytogenes
XX EGD-e. This sequence and fragments of this sequence are useful for
XX selecting probes and primers for detecting genes in L. monocytogenes and
XX related organisms, and to study genetic polymorphisms and other genomes.
XX proteins (ABB47297-ABB50149) expressed from the present sequence are
XX useful for raising specific antibodies, identification of L.
XX monocytogenes and related organisms, and for biosynthesis and
XX biodegradation, especially biosynthesis of Vitamin B12. This sequence and
XX proteins encoded by it are also useful for selecting compounds that
XX regulate gene expression and cell replication and modulate L.
XX monocytogenes-related diseases. In addition, this sequence and proteins
XX encoded by it are useful in pharmaceutical and vaccines compositions for
XX the treatment or prevention of infections by L. monocytogenes and related
XX organisms.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 2944528 BP; 914202 A; 563301 C; 555061 G; 911964 T; 0 other;

Query Match      22.6%; Score 34.6; DB 24; Length 2944528;
Best Local Similarity 59.8%; Pred. NO. 11;
Matches 58; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 6 TCCAGATGATCTGTATATAAAACCAACAGCAACAGTTACAGACCAAAACCAACCAAGC 65
Db 2393914 TTCAAAGTTCCCTAGCAAAATACCCAAAGATAACAGTTATATAATCTAATAATTACAATC 2393855

QY 66 AAAACAGTTACAGACCGCTGCAGACTACCTCATCAAA 102
Db 2393854 AAAATAGTATAGTACGACCAACAAACAAACAACTACA 2393818

Search completed: August 2, 2002, 20:47:17
Job time: 25928 sec
```



Mon Aug 5 11:51:49 2002

us-08-973-363-2.rng

Page 11



GenCore version 4.5  
Copyright (C) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 2, 2002, 16:52:08 ; Search time 8940.62 Seconds  
(without alignments)  
230.972 Million cell updates/sec

Title: US-08-973-363-2  
Perfect score: 153  
Sequence: 1 ATTCCTCAGATGATCTGCA.....CTCAGAGACTTTGTGTGGCG 153

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*  
1: em\_estba.\*  
2: em\_esthum.\*  
3: em\_estlin.\*  
4: em\_estnu.\*  
5: em\_estov.\*  
6: em\_estpi.\*  
7: em\_estro.\*  
8: em\_hic.\*  
9: gb\_esti.\*  
10: gb\_estc.\*  
11: gb\_hic.\*  
12: gb\_gss.\*  
13: em\_gss\_hum.\*  
14: em\_gss\_inv.\*  
15: em\_gss\_pln.\*  
16: em\_gss\_vrt.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	102	66.7	619	9	BB155356
2	102	66.7	660	9	BB461065
3	95.6	62.5	446	9	BB834922
4	89.4	58.4	438	9	BB830730
5	77	50.3	430	9	AL601246
6	77	50.3	547	9	AL890775
7	77	50.3	821	10	BF239967
8	77	50.3	866	9	AU125712
9	77	50.3	1028	10	BE895133
10	72.2	47.2	686	9	AW997058
11	63	41.2	337	9	AW996787
12	56.8	37.1	593	9	AL659353
13	56.8	37.1	645	9	AL644594
14	50.8	33.2	856	12	CNS04DVG
15	42.4	27.7	642	9	AL658477
16	42	27.5	352	12	AQ275532
17	41.6	27.2	424	9	AV546625

18	41.4	27.1	297	10	H53538
19	41.2	26.9	129	10	BF881342
20	41.2	26.9	136	9	AI085785
21	41.2	26.9	300	10	AI09703
22	41.2	26.9	334	9	AN305759
23	41.2	26.9	356	10	H61027
24	41.2	26.9	436	9	AI848536
25	41.2	26.9	446	10	H53564
26	41.2	26.9	546	10	BG757172
27	41.2	26.9	667	9	AW364689
28	41	26.8	366	9	AI915135
29	39.8	26.0	532	10	BI844947
30	39.8	26.0	574	10	BM185888
31	39.8	26.0	578	10	BM070535
32	39.8	26.0	608	10	BM024825
33	39.8	26.0	608	10	BM185005
34	39.6	25.9	339	9	AI870450
35	38.6	25.2	1072	12	CNS0070B
36	37.4	24.4	619	9	BB646029
37	37.2	24.3	1088	10	BF133627
38	37.2	24.1	243	10	AW168274
39	36.8	24.1	243	10	C84397
40	36.8	24.1	290	10	BM362956
41	36.8	24.1	403	10	BM365102
42	36.8	24.1	414	10	BM366198
43	36.8	24.1	417	10	BM365756
44	36.8	24.1	420	10	BM364634
45	36.8	24.1	442	10	BM366212

## ALIGNMENTS

RESULT 1

BB155356  
LOCUS  
DEFINITION  
musculus cDNA clone AL30024L16 3' similar to L10410 Mouse  
DNA-binding protein (CHD-1) mRNA, mRNA sequence.  
EST.  
BB155356  
BB155356.2 GI:16268254  
EST.  
house mouse.  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 619)  
Arakawa, T., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A.,  
Hiramoto, K., Horii, F., Ishii, Y., Ito, M., Kawai, J., Konno, H., Kouda  
, M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, K., Ohno, H., Sasaki  
Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki  
D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H.,  
Tagami, M., Tagawa, A., Takahashi, F., Takeda, Y., Tanaka, T., Toya, T.,  
Muramatsu, M., Hayashizaki, Y.,  
RIKEN Mouse ESTs (Arakawa, T., et al. 2001)  
Unpublished (2001)  
On Jun 29, 2000 this sequence version replaced gi:8811286.  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Tsurumi-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@sc.riken.go.jp/  
URL: http://genome.gsc.riken.go.jp/  
Carninci, P., Shibata, Y., Hayashizaki, Y., Muramatsu, M., Sugahara, Y., Itoh  
, M., Konno, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. Genome Res. 10 (10), 1617-1630 (2000)  
wagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,  
Watabiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura

S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)  
Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y. and Hayashizaki, Y.  
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
Kondo, S., Shingawa, A., Saito, T., Kiyosawa, H., Yamanaka, I., Aizawa, K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K. and Hayashizaki, Y.  
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)  
Please visit our web site (<http://genome.gsc.riken.go.jp/>) for further details.  
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

#### FEATURES

Location/Qualifiers

1..619  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="A130024L16"  
/clone\_lib="RIKEN full-length enriched, 16 days neonate thymus"  
/tissue\_type="thymus"  
/dev\_stage="16 days neonate"  
/lab\_host="DHI0B"  
/note="Site\_1: Sali; Site\_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5'  
GAGAGAGAAGGATCCAGAGCTCTTTTATTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 185.0. Second strand cDNA was prepared with the primer adaptor of sequence [5' GAGAGAGATCTCGACTTAATAATTCCTCCCTCCCTCC 3']. cDNA was cleaved with xhoI and BamHI. Vector: a modified pluescript KS(+) after bulk excision from Lambda

BASE COUNT 194 a 118 c 161 g 146 t  
ORIGIN

Query Match 66.7%; Score 102; DB 9; Length 619;

Best Local Similarity 100.0%; Pred. No. 4e-11;  
Matches 102; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 52 AAAAAACACAAAGCAAGTTCAGACCCCTGCAGACTCCTCATCAACTACTTAGC 111  
|||||

Db 510 AAAAAACACAAAGCAAGTTCAGACCCCTGCAGACTCCTCATCAACTACTTAGC 569  
|||||

QY 112 AGAGATCTTGGCAAAAGAGGCTCAGAGACTTTGTGTGGC 153  
|||||

Db 570 AGAGATCTTGGCAAAAGAGGCTCAGAGACTTTGTGTGGC 611  
|||||

#### RESULT 2

BB461065

LOCUS

BB461065 660 bp mRNA linear EST 25-OCT-2001  
BB461065 RIKEN full-length enriched, 12 days embryo spinal ganglion  
Mus musculus cDNA clone D130070B13 3' similar to L10410 Mouse  
DNA-binding protein (CHD-1) mRNA, mRNA sequence.

#### ACCESSION

BB461065

#### VERSION

KEYWORDS

SOURCE

ORGANISM

BB461065.2 GI:16426612

EST

house mouse.

Mus musculus

REFERENCE

AUTHORS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 660)  
Arakawa, T., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hiramoto, K., Hori, F., Ishii, Y., Ito, M., Kawai, J., Konno, H., Kouda, M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, K., Ohno, M., Okazaki, Y., Okido, F., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shingawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Takeda, Y., Tanaka, T., Toyota, T., Muramatsu, M. and Hayashizaki, Y.  
RIKEN Mouse ESTs Arakawa, T., et al. 2001)  
Unpublished (2001)

TITLE

JOURNAL

COMMENT

On Jul 21, 2000 this sequence version replaced gi:9356558.  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@gsc.riken.go.jp,  
URL: <http://genome.gsc.riken.go.jp/>  
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)  
wagi, K., Fujiwaki, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsumura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)  
Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y. and Hayashizaki, Y.  
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
Kondo, S., Shingawa, A., Saito, T., Kiyosawa, H., Yamanaka, I., Aizawa, K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K. and Hayashizaki, Y.  
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)  
Please visit our web site (<http://genome.gsc.riken.go.jp/>) for further details.  
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.  
Location/Qualifiers  
1..660  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="D130070B13"  
/clone\_lib="RIKEN full-length enriched, 12 days embryo spinal ganglion"  
/tissue\_type="spinal ganglion"  
/dev\_stage="12 days embryo"  
/lab\_host="DHI0B"  
/note="Site\_1: Sali; Site\_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5'  
GAGAGAGAAGCGCCGCACTCGAGTCTTTTATTTTNN 3'], cDNA was

prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5', GAGAGAGAGATCTCGAGTTAATTAATCCGCCGCCCCC 3']. cDNA was cleaved with BamHI and XhoI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."

BASE COUNT  
ORIGIN

211 a 130 c 168 g 151 t

Query Match 66.7%; Score 102; DB 9; Length 660;  
Best Local Similarity 100.0%; Pred. No. 3.9e-11;  
Matches 102; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 52 AAAAAACACAGCAGACAGTACAGACCGTGCAGACTCCTCAACTACTTACG 111  
|||||  
DB 502 AAAAAACACAGCAGACAGTACAGACCGTGCAGACTCCTCAACTACTTACG 561  
|||||

QY 112 AGAGATCTTGCAGAGAGGCTCAGAGACTTTGTGTGCG 153  
|||||

DB 562 AGAGATCTTGCAGAGAGGCTCAGAGACTTTGTGTGCG 603  
|||||

RESULT 3  
BB834922  
LOCUS  
DEFINITION  
Jyg-MC(B) cDNA Mus musculus cDNA clone G930033J21 3', mRNA  
sequence.  
BB834922  
EST.  
SOURCE  
ORGANISM  
house mouse.  
Mus musculus  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
1 (bases 1 to 446)  
Akimura,T., Hirakawa,T., Carninci,P., Furuno,M., Hanagaki,T.,  
Hayatsu,N., Hiramoto,K., Hirooka,T., Hirozane,T., Imotani,K., Ishii  
Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T.,  
Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T.,  
Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K.,  
Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa  
A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T.,  
Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.  
RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.  
2001)

Unpublished (2001)  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@sc.riken.go.jp,  
URL:http://genome.gsc.riken.go.jp/  
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh  
M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. Genome Res. 10 (10), 1617-1630 (2000)  
Wagi,K., Fujiwaka,S., Inoue,K., Togawa,M., Izawa,M., Ohara,E.,  
Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura  
S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and  
Hayashizaki,Y.  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
10 (11), 1757-1771 (2000)  
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara  
Y. and Hayashizaki,Y.  
Computer-based methods for the mouse full-length cDNA

encyclopedia: real-time sequence clustering for construction of a  
non-redundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
Please visit our web site (<http://genome.gsc.riken.go.jp>) for  
further details.

e mouse tissues.

FEATURES  
source

Location/Qualifiers  
1..446  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="G930033J21"  
/clone\_lib="RIKEN full-length enriched, mammary gland  
RCB-0527 Jyg-MC(B) cDNA"  
/tissue\_type="mammary gland"  
/cell\_line="RCB-0527 Jyg-MC(B)"

BASE COUNT 139 a 99 c 108 g 100 t

Query Match 62.5%; Score 95.6; DB 9; Length 446;  
Best Local Similarity 96.1%; Pred. No. 7.9e-10;  
Matches 98; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 52 AAAAAACACAGCAGTACAGACCGTGCAGACTTCTCAACTACTTACG 111  
|||||  
DB 330 AAAAAACACAGCAGTACAGACCGTGCAGACTTCTCAACTACTTACG 389  
|||||

QY 112 AGAGATCTTGCAGAGAGGCTCAGAGACTTTGTGTGCG 153  
|||||

DB 390 AGAGATCTTGCAGAGAGGCTCAGAGACTTTGTGTGCG 431  
|||||

RESULT 4  
BB830730  
LOCUS  
DEFINITION  
Jyg-MC(B) cDNA Mus musculus cDNA clone G930013K04 3', mRNA  
sequence.  
BB830730  
EST.  
SOURCE  
ORGANISM  
house mouse.  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 438)  
Akimura,T., Hirakawa,T., Carninci,P., Furuno,M., Hanagaki,T.,  
Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T.,  
Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T.,  
Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K.,  
Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa  
A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T.,  
Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.  
RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.  
2001)

Unpublished (2001)  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@sc.riken.go.jp,  
URL:http://genome.gsc.riken.go.jp/  
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh  
M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. Genome Res. 10 (10), 1617-1630 (2000)  
Wagi,K., Fujiwaka,S., Inoue,K., Togawa,M., Izawa,M., Ohara,E.,  
Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura  
S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and  
Hayashizaki,Y.

Hayashizaki, Y.  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer. Genome Res. .  
10 (11), 1757-1771 (2000)  
Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara  
, Y. and Hayashizaki, Y.  
Computer-based methods for the mouse full-length cDNA  
encyclopedia: real-time sequence clustering for construction of a  
nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)  
Please visit our web site (<http://genome.gsc.riken.go.jp>) for  
further details.

e mouse tissues.

FEATURES  
source  
Location/Qualifiers  
1. 438  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="G930013K04"  
/clone\_lib="RIKEN full-length enriched, mammary gland  
RCB-0527 Jyg-MC(B) cDNA"  
/tissue\_type="mammary gland"  
/cell\_line="RCB-0527 Jyg-MC(B)"  
BASE COUNT 138 a 96 c 108 g 96 t  
ORIGIN

Query Match 58.4%; Score 89.4; DB 9; Length 438;  
Best Local Similarity 98.1%; Pred. No. 1.3e-08;  
Matches 101; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 52 AAAAAACCAAGCAAAACAGTTACAGACCCGTCGAGACTACCTCATCAAACTA-CTTAG 110  
|||||  
DB 321 AAAAAACCAAGCAAAACAGTTACAGACCCGTCGAGACTACCTCATCAAACTAGCTTAG 380  
|||||

QY 111 CAGAGATCTTGC AAAAGAGGCTCAGAGACTTTGTGTGGG 153  
|||||  
DB 381 CAGAGATCTTGC AAAAGAGAGGCTCAGAGACTTTGTGTGGG 423  
|||||

RESULT 5  
LOCUS AL601246 430 bp mRNA linear EST 14-AUG-2001  
DEFINITION DKFZp313J1040\_r1 313 (synonym: hlcc2) Homo sapiens cDNA clone  
AL601246  
ACCESSION AL601246  
VERSION AL601246.1 GI:15164752  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 430)  
AUTHORS Bloecker, H., Boecher, M., Brandt, P., Mewes, W., Well, B. and Wiemann  
, S.  
TITLE EST (Bloecker, H., Boecher, M., Brandt, P., Mewes, H.W., Well, B. and  
Wiemann, S.)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Bloecker H  
MIPS

Am Klopferspitz 18a D-82152 Martinsried, Germany  
This is the 5' sequence of the clone insert  
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer  
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;  
Sequenced by GBF (National Research Centre for Biotechnology Ltd.,  
Braunschweig/Germany) within the cDNA sequencing consortium of the  
German Genome Project.

No s1 sequence available.  
This clone (DKFZp313J1040) is available at the RZPD in Berlin.  
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059  
Berlin-Charlottenburg, GERMANY; Email: [clone@rzpd.de](mailto:clone@rzpd.de).

FEATURES  
source  
Location/Qualifiers  
1. 430  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"

/clone="DKFZp313J1040"  
/clone\_lib="313 (synonym: hlcc2)"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/note="Vector: pRipEx2; Site\_1: SfiIA; Site\_2: SfiIB;  
cDNA-collection"  
BASE COUNT 163 a 81 c 86 g 100 t  
ORIGIN

Query Match 50.3%; Score 77; DB 9; Length 430;  
Best Local Similarity 94.1%; Pred. No. 3.8e-06;  
Matches 80; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 52 AAAAAACCAAGCAAAACAGTTACAGACCCGTCGAGACTACCTCATCAAACTACTAGC 111  
|||||  
DB 239 AAAAAACCAAGCAAAACAGTTACAGACCCGTCGAGACTACCTCATCAAACTACTAGT 298  
|||||  
QY 112 AGAGATCTTGC AAAAAGAGAGGCTC 136  
|||||  
DB 299 AGAGATCTTGC AAAAAGAGAGGCTC 323  
|||||

RESULT 6  
LOCUS AI890775/c 547 bp mRNA linear EST 07-MAR-2000  
DEFINITION WM95f11.x1 NCI-CGAP Ut2 Homo sapiens cDNA clone IMAGE:2443725 3'  
similar to SW:CHDL\_HUMAN 014646 CHROMODOMAIN-HELICASE-DNA-BINDING  
PROTEIN 1 ; mRNA sequence.  
ACCESSION AI890775  
VERSION AI890775.1 GI:5595939  
KEYWORDS EST.  
SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 547)  
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index

JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: Life Technologies, Inc.  
DNA Sequencing by: Greg Lennon, Ph.D.  
Clone Distribution: Washington University Genome Sequencing Center  
found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)  
Insert Length: 1924 Std Error: 0.00  
Seq primer: -40UP from Gibco  
High quality sequence stop: 418.

FEATURES  
source  
Location/Qualifiers  
1. 547  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2443725"  
/clone\_lib="NCI-CGAP\_Ut2"  
/tissue\_type="moderately-differentiated endometrial  
adenocarcinoma, 3 pooled tumors"  
/lab\_host="DH10B"  
/note="Organ: uterus; Vector: pCMV-SPORT6; Site\_1: SalI;  
Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dt.  
Average insert size 1.85 kb. Life Technologies catalog #:  
11539-012"  
BASE COUNT 114 a 118 c 85 g 230 t  
ORIGIN

Query Match 50.3%; Score 77; DB 9; Length 547;  
Best Local Similarity 94.1%; Pred. No. 3.6e-06;



**TITLE** National Institutes of Health, Mammalian Gene Collection (MGC)  
**JOURNAL** Unpublished (1999)  
**COMMENT** Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: ATCC/DCTD/DTF

cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LIA9753 row: h column: 16  
High quality sequence stop: 488.

**FEATURES**  
source  
1. .1028  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3921087"  
/clone\_lib="NIH\_MGC\_72"  
/tissue\_type="melanotic melanoma"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: skin; Vector: pCMV-SPORT6; Site\_1: NotI; Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 2 kb. Library constructed by Life Technologies."  
387 a 205 c 238 g 198 t

**BASE COUNT**  
ORIGIN

**Query Match** 50.3%; Score 77; DB 10; Length 1028;  
**Best Local Similarity** 94.1%; Pred. No. 3.2e-06;  
**Matches** 80; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

**Qy** 52 AAAAAACACAGCAAAACAGTTACAGACCGTGCAGACTCCTCATCAAACTACTAGC 111  
|||||  
**Db** 152 AAAAAACACAGCAAAACAGTTACAGACCGTGCAGACTCCTCATCAAACTACTAGT 211  
|||||

**Qy** 112 AGAGATCTTGCAAAAAGAGAGGCTC 136  
|||||

**Db** 212 AGAGATCTTGCAAAAAGAGAGCTC 236  
|||||

**RESULT** 10  
AW997058/c  
**LOCUS** AW997058 686 bp mRNA linear EST 05-JUN-2000  
**DEFINITION** QV3-BN0047-150400-152-c03 BN0047 Homo sapiens cDNA, mRNA sequence.  
**ACCESSION** AW997058  
**VERSION** AW997058.1 GI:8257292  
**KEYWORDS** EST.  
**SOURCE** human.

**ORGANISM** Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
**REFERENCE** 1 (bases 1 to 686)  
**AUTHORS** Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.

**TITLE** Shotgun sequencing of the human transcriptome with ORF expressed sequence tags  
**JOURNAL** Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
**MEDLINE** 20202663  
**COMMENT** Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome

Project. This entry can be seen in the following URL  
(<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=st2-QV3-BN0047-150400-152-c03&t3=2000-04-15&t4=1>)  
**Seq primer:** puc 18 forward  
**High quality sequence start:** 19  
**High quality sequence stop:** 678.

**FEATURES**  
source

1. .686  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="BN0047"  
/dev\_stage="Adult"  
/note="Organ: breast\_normal; Vector: puc18; Site\_1: SmaI; Site\_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."  
165 a 154 c 126 g 241 t

**BASE COUNT**  
ORIGIN

**Query Match** 47.2%; Score 72.2; DB 9; Length 686;  
**Best Local Similarity** 96.1%; Pred. No. 3.1e-05;  
**Matches** 74; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

**Qy** 53 AAAACACAGCAAAACAGTTACAGACCGTGCAGACTCCTCATCAAACTACTAGCA 112  
|||||

**Db** 106 AAAACACCAAGCAAAACAGTTACAGACCGTGCAGACTCCTCATCAAACTACTAGTA 47  
|||||

**Qy** 113 GAGATCTTGCAAAAAGA 129  
|||||

**Db** 46 GAGATCTTGCAAAAAGA 30  
|||||

**RESULT** 11  
AW996787/c

**LOCUS** AW996787 337 bp mRNA linear EST 05-JUN-2000  
**DEFINITION** QV3-BN0047-230200-102-d03 BN0047 Homo sapiens cDNA, mRNA sequence.  
**ACCESSION** AW996787  
**VERSION** AW996787.1 GI:8257021  
**KEYWORDS** EST.  
**SOURCE** human.

**ORGANISM** Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
**REFERENCE** 1 (bases 1 to 337)  
**AUTHORS** Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.

**TITLE** Shotgun sequencing of the human transcriptome with ORF expressed sequence tags  
**JOURNAL** Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
**MEDLINE** 20202663  
**COMMENT** Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
(<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=st2-QV3-BN0047-230200-102-d03&t3=2000-02-23&t4=1>)  
**Seq primer:** puc 18 forward  
**High quality sequence start:** 2  
**High quality sequence stop:** 337.



FEATURES  
source

Location/Qualifiers  
1. .337  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="BN0047"  
/dev\_stage="Adult"  
/note="Organ: breast\_normal; Vector: puc18; Site\_1: SmaI;  
Site\_2: SmaI; A mini-library was made by cloning products  
derived from ORESTES PCR (U.S. Letters Patent application  
No. 196,716 - Ludwig Institute for Cancer Research)  
profiles into the pUC 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under  
low stringency conditions."  
83 a 59 c 72 g 123 t

BASE COUNT  
ORIGIN

Query Match 41.2%; Score 63; DB 9; Length 337;  
Best Local Similarity 91.8%; Pred. No. 0.0024;  
Matches 78; Conservative 0; Mismatches 5; Indels 2; Gaps 1;

QY 52 AAAAACCACAGCAAAACAGTTACAGACCGGTGCAGACTACCTCATCAACTACTTACG 111  
|||||  
Db 84 AAAAACCACAGCAAAACAGTTACAGACCGGTGCAGACTACCTCATCAACTACTTACG 111  
|||||

QY 112 AGAGATCTTGCAGAAAGAGAGGCTC 136  
|||||

Db 26 AGAGATCTTGCAGAAAGAGAGGCTC 2

RESULT 12  
AL659353/c  
LOCUS  
DEFINITION AL659353 XGC-neurula Silurana tropicalis cDNA clone TNeu045e20 5',  
mRNA sequence.  
ACCESSION AL659353  
VERSION AL659353.1 GI:17672995  
KEYWORDS EST.  
SOURCE Western clawed frog.  
ORGANISM Silurana tropicalis  
REFERENCE  
AUTHORS Huckle,E., Taylor,R., Ashurst,J.L., Zorn,A.M. and Rogers,J.  
TITLE Sanger Xenopus tropicalis EST project 2001 (10\_2001)  
JOURNAL Unpublished (2001)  
COMMENT Contact: Huckle E  
Sanger Centre  
Hinxton, Cambridgeshire, CB10 1SA, UK  
Email: trop@sanger.ac.uk  
Sanger Xenopus tropicalis EST project 2001  
TROPICALIS\_SEQUENCE\_ID: TNeu045e20.sp6  
Sequencing primer: SP6  
This sequence is from a Xenopus Gene Collection (XGC) library  
constructed by Aaron M. Zorn.  
Location/Qualifiers  
1. .593  
/organism="Silurana tropicalis"  
/db\_xref="taxon:8364"  
/clone="TNeu045e20"  
/clone\_lib="XGC-neurula"  
/dev\_stage="neurula"  
/lab\_host="Escherichia coli DH10B"  
/note="Vector: pCS107; Site\_1: EcoRI; Site\_2: NotI; cDNA  
was oligo dT primed from 5ug of poly A+ RNA from neurula.  
EcoRI-NotI cut cDNA was then ligated into pCS107 with  
EcoRI at the 5' end and NotI at the 3' end."  
119 a 134 c 114 g 226 t

BASE COUNT  
ORIGIN

Query Match 37.1%; Score 56.8; DB 9; Length 593;

FEATURES  
source

Location/Qualifiers  
1. .645  
/organism="Silurana tropicalis"  
/db\_xref="taxon:8364"  
/clone="L1E1d12"  
/clone\_lib="XGC-egg"  
/dev\_stage="egg"  
/lab\_host="Escherichia coli XL1-blue"  
/note="vector: pCS107; Site\_1: EcoRI; Site\_2: NotI; cDNA  
was oligo dT primed from 5ug of poly A+ RNA from egg.  
EcoRI-NotI cut cDNA was then ligated into pCS107 with  
EcoRI at the 5' end and NotI at the 3' end"  
222 a 125 c 156 g 141 t 1 others

BASE COUNT  
ORIGIN

Query Match 37.1%; Score 56.8; DB 9; Length 645;  
Best Local Similarity 73.0%; Pred. No. 0.035;  
Matches 73; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

QY 52 AAAAACCACAGCAAAACAGTTACAGACCGGTGCAGACTACCTCATCAACTACTTACG 111  
|||||  
Db 477 AAGAAGCCCAAGCCAGCAGCTACAGACCGGTGCAGACTACCTCATCAACTACTTCAAT 536  
|||||

QY 112 AGAGATCTTGCAGAAAGAGAGGCTCAGAGCTTTGTGGTG 151  
|||||

Db 537 AAGAATCTGCTTAGGAAAGAGCAACAAAGACTTTCTTAGTG 576  
|||||

RESULT 14  
CNS04DVG/c  
LOCUS  
DEFINITION CNS04DVG 856 bp DNA linear GSS 21-MAY-2000  
Tetraodon nigroviridis genome survey sequence T7 end of clone  
103K08 of library G from Tetraodon nigroviridis, genomic survey  
sequence.



Mon Aug 5 11:51:49 2002

us-08-973-363-2.1st

Page 9

**THIS PAGE BLANK (USPTO)**



```
Query Match      100.0%; Score 153; DB 6; Length 153;
Best Local Similarity 100.0%; Pred. No. 2.2e-35;
Matches 153; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATTTTACCTGATGATCCAGACAAAGAAACCCAGGCAAGCAGCTACAGACCAAGAAACCC 60
|||||
Db 1 ATTTTACCTGATGATCCAGACAAAGAAACCCAGGCAAGCAGCTACAGACCAAGAAACCC 60
|||||

QY 61 CAGGCAAAAGCAGCTACAGACCCGTCGAGACTACCTCATTAATTACTGAATAAGACCTT 120
|||||
Db 61 CAGGCAAAAGCAGCTACAGACCCGTCGAGACTACCTCATTAATTACTGAATAAGACCTT 120
|||||

QY 121 GCAAGAAAGGAAGCAAAAGGCTTGCTGGTGCA 153
|||||
Db 121 GCAAGAAAGGAAGCAAAAGGCTTGCTGGTGCA 153
|||||

RESULT 2
A58685
LOCUS A58685 153 bp DNA linear PAT 06-MAR-1998
DEFINITION Sequence 4 from Patent WO9639505.
ACCESSION A58685
VERSION A58685.1 GI:3714248
KEYWORDS
SOURCE unidentified.
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 153)
AUTHORS Griffiths,R. and Tiwari,B.
TITLE AVIAN GHD GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN BIRDS
JOURNAL
COMMENT Patent: WO 9639505-A 4 12-DEC-1996;
OTHER PUBLICATION AU 5906996 961224.
FEATURES
source
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 56 a 36 c 31 g 30 t
ORIGIN

Query Match      91.6%; Score 140.2; DB 6; Length 153;
Best Local Similarity 94.8%; Pred. No. 1.5e-31;
Matches 145; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 ATTTTACCTGATGATCCAGACAAAGAAACCCAGGCAAGCAGCTACAGACCAAGAAACCC 60
|||||
Db 1 ATTTTACCTGATGATCCAGACAAAGAAACCCAGGCTAAGCAGTAAAGCAGTACAGACCAAGAAACCC 60
|||||

QY 61 CAGGCAAAAGCAGCTACAGACCCGTCGAGACTACCTCATTAATTACTGAATAAGACCTT 120
|||||
Db 61 CAGGCTAAGCAGTACAGACCCGTCGAGACTACCTCATTAATTACTGAATAAGACCTT 120
|||||

QY 121 GCAAGAAAGGAAGCAAAAGGCTTGCTGGTGCA 153
|||||
Db 121 GCAAGAAAGGAAGCAGACAGACTTGCTGGTGCA 153
|||||

RESULT 3
A58686
LOCUS A58686 153 bp DNA linear PAT 06-MAR-1998
DEFINITION Sequence 5 from Patent WO9639505.
ACCESSION A58686
VERSION A58686.1 GI:3714249
KEYWORDS
SOURCE unidentified.
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 153)
AUTHORS Griffiths,R. and Tiwari,B.

Query Match      85.4%; Score 130.6; DB 6; Length 153;
Best Local Similarity 90.8%; Pred. No. 1.1e-28;
Matches 139; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 1 ATTTTACCTGATGATCCAGACAAAGAAACCCAGGCAAGCAGCTACAGACCAAGAAACCC 60
|||||
Db 1 ATTTTACCTGATGATCCAGATAGAAACCCACAGCAAGCAGTTGCAGACCAAGAAACCA 60
|||||

QY 61 CAGGCAAAAGCAGCTACAGACCCGTCGAGACTACCTCATTAATTACTGAATAAGACCTT 120
|||||
Db 61 CAGGCAAAAGCAGCTACAGACCCGTCGAGACTACCTCATTAATTACTGAATAAGACCTT 120
|||||

QY 121 GCAAGAAAGGAAGCAAAAGGCTTGCTGGTGCA 153
|||||
Db 121 GCAAGAAAGGAAGTGCARAGACTTACTGGTGCA 153
|||||

RESULT 4
A58691
LOCUS A58691 6608 bp DNA linear PAT 06-MAR-1998
DEFINITION Sequence 10 from Patent WO9639505.
ACCESSION A58691
VERSION A58691.1 GI:3714250
KEYWORDS
SOURCE unidentified.
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 6608)
AUTHORS Griffiths,R. and Tiwari,B.
TITLE AVIAN GHD GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN BIRDS
JOURNAL
COMMENT Patent: WO 9639505-A 10 12-DEC-1996;
OTHER PUBLICATION AU 5906996 961224.
FEATURES
source
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 2289 a 1207 c 1459 g 1653 t
ORIGIN

Query Match      67.3%; Score 103; DB 6; Length 6608;
Best Local Similarity 100.0%; Pred. No. 2.1e-20;
Matches 103; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 51 CAAGAAACCCAGGCAAGCAGCTACAGACCCGTCGAGACTACCTCATTAATTACTGAA 110
|||||
Db 4100 CAAGAAACCCAGGCAAGCAGCTACAGACCCGTCGAGACTACCTCATTAATTACTGAA 4159
|||||

QY 111 TAAAGACCTTGCAAGAAAGGAAGCACAAGGCTTGCTGGTGCA 153
|||||
Db 4160 TAAAGACCTTGCAAGAAAGGAAGCACAAGGCTTGCTGGTGCA 4202
|||||

RESULT 5
AF004397
LOCUS AF004397 6872 bp mRNA linear VRT 08-OCT-1997
DEFINITION Gallus gallus chromo-helicase-DNA-binding on the Z chromosome protein, variant with hydrophilic domain, (CHD-Z) mRNA, complete
```

Query Match	67.3%;	Score 103;	DB 5;	Length 6872;
Best Local Similarity	100.0%;	Pred. No. 2.1e-20;		
Matches 103;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	51	CAAGAAACCCGACGCAAGACGCTACAGACCCCGTCGACGACTACCTTAATTAATTAAGTAA 110		
Db	4100	CAAGAAACCCGACGCAAGACGCTACAGACCCCGTCGACGACTACCTTAATTAATTAAGTAA 4159		
QY	111	TAAAGACCTTGCAAGAAGGAAGACACAAAGCGTTGCTGGTGCA 153		
Db	4160	TAAAGACCTTGCAAGAAGGAAGACACAAAGCGTTGCTGGTGCA 4202		
RESULT	6			
LOCUS	A58683	A58683	153 bp	DNA linear PAT 06-MAR-1998
DEFINITION	Sequence 2 from Patent W09639505.			
ACCESSION	A58683			
VERSION	A58683.1	GI:3714246		
KEYWORDS	.	unidentified.		
SOURCE	ORGANISM	unidentified.		
REFERENCE	1 (bases 1 to 153)			
AUTHORS	Griffiths R. and Tiwari, B.			
TITLE	AVIAN GHG GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN BIRDS			
JOURNAL	Patent: WO 9639505-A 2 12-DEC-1996;			
COMMENT	ISIS INNOVATION (GB)			
FEATURES	Other publication AU 5906996 961224.			
source	1..153	Location/Qualifiers		
	/organism="unidentified"			
	/db_xref="taxon:32644"			
BASE COUNT	61 a 37 c 27 g	28 t		
ORIGIN				
Query Match	66.9%;	Score 102.4;	DB 6;	Length 153;
Best Local Similarity	79.6%;	Pred. No. 3e-20;		
Matches 121;	Conservative 0;	Mismatches 31;	Indels 0;	Gaps 0;
QY	1	ATTTTACCTGATGATCCAGACAAGAAACCCCGGCAAAAGCAGCTACAGACCAAGAAACCC 60		
Db	1	ATTTCTCCAGATGATCCTGATAAAAAGACAAGCAAAACAGTGTACAGACCAAAAAACCA 60		
QY	61	CAGGCAAGACGCTACAGACCCGTCGAGACTACCTCATTAATTAATTAAGACCTT 120		
Db	61	CAAGCAAAACAGTTTACAGACCCGTCGAGACTACCTCATCAAACTACTTTAGCAGAGATCTT 120		
QY	121	GCAAGAAAGGAAGCACAAAGCGTTGCTGGTGC 152		
Db	121	GCAAAAGAGGCTCAGAGACTTTGTGTGC 152		
RESULT	7			
LOCUS	A58696	A58696	1316 bp	DNA linear PAT 06-MAR-1998
DEFINITION	Sequence 15 from Patent W09639505.			
ACCESSION	A58696			
VERSION	A58696.1	GI:3714253		
KEYWORDS	.	unidentified.		
SOURCE	ORGANISM	unidentified.		
REFERENCE	1 (bases 1 to 1316)			
AUTHORS	Griffiths, R. and Tiwari, B.			
TITLE	AVIAN GHG GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN BIRDS			
JOURNAL	Patent: WO 9639505-A 15 12-DEC-1996;			
COMMENT	ISIS INNOVATION (GB)			
FEATURES	Other publication AU 5906996 961224.			
Location/Qualifiers				

```
source 1..1316
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 493 a 205 c 308 g 304 t 6 others
ORIGIN

Query Match 60.4%; Score 92.4; DB 6; Length 1316;
Best Local Similarity 94.1%; Pred. No. 3e-17;
Matches 96; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 52 AGAATCCAGGCAAGAGCTACAGACCCGTCGAGACTACCTCATTAATTAAGTGAAT 111
|||||
Db 1159 AAGAATCCAGGCTAAGCAGTACAGACTCGTCGAGATTACCTCATTAATTAAGTGAAT 1218
|||||

QY 112 AAGACCTTGCAGAAAGCAACAAAGGCTTGCTGGTGCA 153
|||||
Db 1219 AAGACCTTGCAGAAAGCAACAAAGGCTTGCTGGTGCA 1260
|||||

RESULT 8
LOCUS D14316 2292 bp mRNA linear VRT 03-FEB-1999
DEFINITION delta-crystallin enhancer binding protein, complete sequence.
ACCESSION D14316
VERSION D14316.1 GI:391639
KEYWORDS ORF2.
SOURCE Gallus gallus (library: lambda gtl1) 13 day embryo lens cdna to
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
1 (bases 1 to 2292)
Funahashi,J.
Direct Submission
Submitted (29-JAN-1993) Jun-ichi Funahashi, Inst. Devel., Aging and
Cancer, Tohoku Univ.; 4-1 Seiryomachi, Aoba-ku, Sendai 980-77,
Japan (Tel:022-272-9499, Fax:022-272-3982)
2 (sites)
Funahashi,J., Sekido,R., Murai,K., Kamachi,Y. and Kondoh,H.
Delta-crystallin enhancer binding protein delta Efi is a zinc
finger-homeodomain protein implicated in postgastrulation
embryogenesis
Development 119 (2), 433-446 (1993)
94116444
3 (bases 1 to 2292)
Funahashi,J.
Unpublished (1994)
Location/Qualifiers
1..2292
/organism="Gallus gallus"
/db_xref="taxon:9031"
/tissue_type="lens"
/clone_lib="lambda gtl1"
/dev_stages="13 day embryo"
257..1939
/note="beta-gal fusion protein binds to blocks 10 to 3 of
HN fragment of delta-crystallin enhancer."
/codon_start=1
/product="ORF2"
/protein_id="BAA03262.1"
/db_xref="GI:391640"
translation="MELKKCCNHCYLIKPPDNEFYNKOEALQHLIRSSGKLILDLK
LIRLRGNRVLIIFSQVMRLDILAEYLKYQFPFQRLDGSIKGELBKALDHFNAEG
SEDFCLLSTPAGSLGILNASADTVIFDSQVNDQNLQAARAHRIQCKQVNIYRL
VTGSEVEDILRAKKKMWLDHLVIQRMDFTKTLHTGTSPPSSTPFNKELSAILK
FGBELPEPEGEQEPQEMDILKRAETRENEPGPLTVGDELLSQFKVANFSNMD
EDDLELPERNRWEELIPESQRRIEEEERQKELEIYMLPRMNCAGKQISFNGE
GRRSRYSQSDSITERRPKKRGPRPTIPRENTKGFSDAIEIRFKYKFGGP
LERLDVARDALVDKSETDLRLGLGVHNCIKALKDNSSGQFAGRGRLGKVGPTF
RISYQVNAKLIVSHEELAPLHKSIPSDPERKRYVYPCHTKAAHFDIDWGKEDDN
LLVGIYAYGSWOCKYVNSRSLNTEILPDPDPTPRQNSIRYPVQTSLNY"

CDS
source
1..5349
/organism="Mus musculus"
/db_xref="taxon:10090"
/cell_type="plasmacytoma"
1..5349
/gene="CHD-1"
171..5306
/gene="CHD-1"
/codon_start=1
/product="DNA-binding protein"
/protein_id="AAB08486.1"
/db_xref="GI:455015"
translation="MNGHSDEESVNRNGSGSSQDGGSGASGSGSGSSSSSSSSDGS
SQSGSDSDSGSDSGSSESDTSRENKVOAKPKPKVDGAEFKWSPSILAVQSRAML
RKQQAQQORPASPSSNGSSESDSDSDSGAKKHKHNDQWMSGSPSQLG
SDSESEERDKSCDGTESDYEPKNVRKPNRKSNGKKTILGOKKROIDSEDE
DDEYDNDKSSRRQATVNVSYKDEEMKTDSDLLLEVCGDVPQPEPERETLERV
DCVRKGATGATTTIYAEADGDPNAGFERNKEPGDIQYLKKWGHSHIINTWETEE
TLQONVNGMKKLDYKKQDQETRWLNKNSPDEVYINCOOELTDDHUKQIIVERI
IAHSNKSAGLPDYCYKQGLPYSECWEDGALISKFKOTCIDYFSRNSKTPFK
DCKVLKQRPREVALKQPSYTGHEGLELDYQLNGLNLAHSCWKNCSCLADBMGL
GKTOTISFLNLFHEQLYGPFLVLPSTLTQWREIQTWASOMNAVVLGDSNR
NMTIHWEMHPQTKRLKFNILLTVEILLDKAFGLGNWAFIGVDEAHRKNDL
XYLLDPSKNSHLLITGTPLONSLKEWLSLLHFIMPEKFSWEEFEHKGREGYGA
SLHLEDFLLRRVKKDVSLPAKVEIOLRWMSALQKQYIKMLITRNYKALSKGSK
GSTGFLLMMELKKKNCILYKPPDNEFYNKOEALQHLIRSSGKLILDLKLR
RERNRVLIQSVMRLDILAEYLKYQFPFQRLDGSIKGELKQALDHFNAEGSEDF
CFLTSRAGGLGINASADTVIFDSQVNDQNLQAARAHRIQCKQVNIYRLVTKG
SVREDILERAKKMWLDHLVIQRMDFTKTLHTGTSPPSSTPFNKELSAILKFGAE
ELFPEGEQEPQEMDILKRAETRENEPGPLTVGDELLSQFKVANFSNMD
ELFPERNSNWEELIPESQRRIEEEERQKELEIYMLPRMNCAGKQISFNGS
RSRYSQSDSITERRPKKRGPRPTIPRENTKGFSDAIEIRFKYKFGGP
ELFPERNSNWEELIPESQRRIEEEERQKELEIYMLPRMNCAGKQISFNGS
RSRYSQSDSITERRPKKRGPRPTIPRENTKGFSDAIEIRFKYKFGGP
```



[illegible]

	Query Match	44.1%	Score 67.4	DB 10	Length 5349
	Best Local Similarity	79.2%	Pred. NO. 9.3e-10		
	Matches 80	Conservative	Mismatches 21	Indels 0	Gaps 0
QY	52	AGAAACCCCGGCAAGCAGCTACAGACCCGTCGAGACTACCTATTAAATTAATCAAT	111		
DB	4047	AAAAACCAACGCAAAACGTTACAGACCCGTCGAGACTACCTATCAAACTACTTAGC	4106		
QY	112	AAAGACCTTCGAAGGAAGGACACAAAGGCTTGTGTGTC	152		
DB	4107	AGAGATCTTCGAAAAGAGAGGCTCAGAGACTTTGTGTGTC	4147		

RESULT					PRI 27-NOV-1997
10					
AF006513					
LOCUS	AF006513			5947 bp	mRNA
DEFINITION	Homo sapiens CHD1 mRNA,			complete cds.	
ACCESSION	AF006513				
VERSION	AF006513.1			GI:2645428	
KEYWORDS					
SOURCE	human				
ORGANISM	Homo sapiens				

REFERENCE  
1 (bases 1 to 5947)  
AUTHORS  
TITLE  
Wooledge, T., Basrai, M.A., Baxeavanis, A.D., Hietter, P. and Collins, F.S.  
Characterization of the CHD family of proteins  
JOURNAL  
Proc. Natl. Acad. Sci. U.S.A. 94 (21), 11472-11477 (1997)  
MEDLINE  
97470991  
REFERENCE  
2 (bases 1 to 5947)

**FEATURES**

Submitted (03-JUN-1997) Laboratory of Gene Transfer, National Human Genome Research Institute, National Institutes of Health, 49 Convent Drive, Bethesda, MD 20892-4442, USA

**Location/Qualifiers**

Direct Submission

```

source
1. .5947
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/map="5q15-21; near WI-5811"
1. .5947
/gene="CHD1"
164. .5293
/gene="CHD1"
/codon_start=1
/product="CHD1"
/protein_id="AAB87381.1"
/db_xref="GI:2645429"

```

[illegible][illegible]

	Query Match	Best Local Similarity	41.0%;	Score 62.8;	DB 9;	Length 5947;
	Matches	76;	Conservative	0;	Mismatches 22;	Indels 0;
QY	52	AGAAACCCAGGCAAGCAGCTACAGACCCGCTGCAGACTACCTCTATAATTTACTGAAT	111			
DB	4046	AAAAACACCAAGCAAAACAGTTCGACCCGCTGCAGACTACCTCATCAAAATTTACTTACT	4105			
QY	112	AAAGACCTTGCAGAAAGCAAGCAAGGCTTTGCTGG	149			
DB	4106	AGAGATCTTGCAGAAAGCAAGGCTTTCTTGTGGTGGGG	4143			

RESULT	11
AC092372/C	
LOCUS	101220 bp DNA linear PRI 07-DEC-2001
DEFINITION	Homo sapiens chromosome 5 clone RP11-58M12, complete sequence.
ACCESSION	AC092372
VERSION	AC092372.3 GI:17402768
KIRBORNS	HTG.
SOURCES	Bombay human
ORGANISM	Homo sapiens

REFERENCE	Eukaryota: Chordata: Craniata: Vertebrata: Euteleostomi:
AUTHORS	Mammalia: Eutheria: Primates; Catarrhini; Hominidae; Homo.
JOURNAL	1 (bases 1 to 101220)
TITLE	DOE Joint Genome Institute and Stanford Human Genome Center.
REFERENCE	Direct Submission
AUTHORS	2 (bases 1 to 101220)
JOURNAL	DOE Joint Genome Institute.
TITLE	Direct Submission
REFERENCE	Submitted (03-JUL-2001) Production Sequencing Facility, DOE Joint
AUTHORS	Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
JOURNAL	3 (bases 1 to 101220)
TITLE	DOE Joint Genome Institute and Stanford Human Genome Center.
REFERENCE	Direct Submission
AUTHORS	Submitted (07-DEC-2001) DOE Joint Genome Institute, 2800 Mitchell
JOURNAL	Drive, Walnut Creek, CA 94598, USA
COMMENT	On Dec 7, 2001 this sequence version replaced qi:15290448.

Draft Sequence Produced by DOE Joint Genome Institute  
[www.jgi.doe.gov](http://www.jgi.doe.gov)  
 Finishing Completed at Stanford Human Genome Center  
[www.shgc.stanford.edu](http://www.shgc.stanford.edu)  
 Quality: Phrap Quality >=40 100% of Sequence;  
 Estimated Total Number of Errors is 0.  
 NOTE: This insert is not the entire sequence of the clone (entire  
 sequence is 146.7kb). It is clipped at the overlap with AC012624.  
 The number of bases overlapped is 90404.  
 Location/Qualifiers

[illegible]

```

* 38821 38920: gap of 100 bp
* 38921 40411: contig of 1491 bp in length
* 40412 40511: gap of 100 bp
* 40512 43279: contig of 2768 bp in length
* 43280 43379: gap of 100 bp
* 43380 46905: contig of 3526 bp in length
* 46906 47005: gap of 100 bp
* 47006 51830: contig of 4825 bp in length
* 51831 51930: gap of 100 bp
* 51931 62619: contig of 10689 bp in length
* 62620 62719: gap of 100 bp
* 62720 75408: contig of 12689 bp in length
* 75409 75508: gap of 100 bp
* 75509 92516: contig of 17008 bp in length
* 92517 92616: gap of 100 bp
* 92617 106409: contig of 13793 bp in length
* 106410 106509: gap of 100 bp
* 106510 143079: contig of 36570 bp in length.
FEATURES
    source
        1..143079
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone_lib="RPC1-II Human Male BAC"
        1..38820
            /note="assembly_fragment"
            vector_side:left
        38921..40411
            /note="assembly_fragment"
        40512..43279
            /note="assembly_fragment"
        43380..46905
            /note="assembly_fragment"
        47006..51830
            /note="assembly_fragment"
        51931..62619
            /note="assembly_fragment"
        62720..75408
            /note="assembly_fragment"
        75509..92516
            /note="assembly_fragment"
        92617..106409
            /note="assembly_fragment"
        106510..143079
            /note="assembly_fragment"
            clone_end:right
            vector_side:right
BASE COUNT 43971 a 26246 c 26678 g 45278 t 906 others
ORIGIN

Query Match 40.9%; Score 62.6; DB 2; Length 143079;
Best Local Similarity 83.5%; Pred. No. 2.7e-08;
Matches 71; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 52 AAGAACCCCGAGCAAGCTACAGACCCGTCAGACTACCTCATTAATTAATTAAT 111
Db 116168 AAAAAACCAAGCAAAACAGTTGCGAGACCCGTCAGACTACCTCATTAATTAATTAAT 116227

QY 112 AAGACCTTGCAGAAAGGAGCAC 136
Db 116228 AGAGATCTTGCAAAAAGAGAGCTC 116252

RESULT 14
AC008531
LOCUS
DEFINITION Homo sapiens chromosome 5 clone CTC-480B11, WORKING DRAFT SEQUENCE,
7 ordered pieces.
AC008531
AC008531.3 GI:128330078
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT.

```

```

SOURCE
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 145659)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 5
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 145659)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT On Feb 14, 2001 this sequence version replaced gi:7528342.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov

Project Information
Center Project Name: 369535
Center clone name: CIR-HSPC_480B11
-----
Summary Statistics
Consensus quality: 139128 bases at least Q40
Consensus quality: 142556 bases at least Q30
Consensus quality: 143744 bases at least Q20
Estimated insert size: 148000; pulse field gel estimation
Quality coverage: 6.27 in Q20 bases; sum-of-contigs estimation
Quality coverage: 6.4 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 56174: contig of 56174 bp in length
* 56175 56274: gap of unknown length
* 56275 100874: contig of 44600 bp in length
* 100875 100974: gap of unknown length
* 100975 113127: contig of 12153 bp in length
* 113128 113227: gap of unknown length
* 113228 118190: contig of 4963 bp in length
* 118191 118290: gap of unknown length
* 118291 119694: contig of 1404 bp in length
* 119695 119795: gap of unknown length
* 119795 123297: contig of 3503 bp in length
* 123298 123397: gap of unknown length
* 123398 145659: contig of 22262 bp in length.
FEATURES
    source
        1..145659
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="5"
            /clone_lib="CalTech human BAC library C"
            /clone_id="CTC-480B11"
BASE COUNT 42561 a 26309 c 27580 g 48609 t 600 others
ORIGIN

Query Match 40.9%; Score 62.6; DB 2; Length 145659;
Best Local Similarity 83.5%; Pred. No. 2.7e-08;
Matches 71; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 52 AAGAACCCCGAGCAAGCTACAGACCCGTCAGACTACCTCATTAATTAATTAAT 111
Db 46791 AAAAAACCAAGCAAAACAGTTGCGAGACCCGTCAGACTACCTCATTAATTAATTAAT 46850

QY 112 AAGACCTTGCAGAAAGGAGCAC 136

```

```
DB 46851 ACAGATCTTGCACCAAGAGAGCTC 46875
|||||
RESULT 15
AC091946 193446 bp DNA linear HTG 09-JUN-2001
LOCUS Homo sapiens chromosome 5 clone RP11-360I2, *** SEQUENCING IN
DEFINITION PROGRESS ***, 33 unordered pieces.
ACCESSION AC091946
VERSION AC091946.1 GI:14333882
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 193446)
REFERENCE DOE Joint Genome Institute.
AUTHORS Sequencing of Human Chromosome 5
TITLE Unpublished
JOURNAL
REFERENCE 2 (bases 1 to 193446)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (09-JUN-2001) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 544799
Center clone name: RPCI-11_360I2
-----
Summary Statistics
Consensus quality: 157767 bases at least Q40
Consensus quality: 180259 bases at least Q30
Consensus quality: 184175 bases at least Q20
Estimated insert size: 204590; agarose-fp estimation
Estimated insert size: 190246; sum-of-contigs estimation
Quality coverage: 5.38 in Q20 bases; agarose-fp estimation
Quality coverage: 5.79 in Q20 bases; sum-of-contigs estimation
NOTE: This is a 'working draft' sequence. It currently
* consists of 33 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1358: contig of 1358 bp in length
* 1359 1458: gap of unknown length
* 1459 2667: contig of 1209 bp in length
* 2668 2767: gap of unknown length
* 2768 3823: contig of 1056 bp in length
* 3824 3923: gap of unknown length
* 3924 5920: contig of 1997 bp in length
* 5921 6020: gap of unknown length
* 6021 7494: contig of 1474 bp in length
* 7495 7594: gap of unknown length
* 7595 9097: contig of 1502 bp in length
* 9097 9196: gap of unknown length
* 9197 10260: contig of 1064 bp in length
* 10261 10360: gap of unknown length
* 10361 12460: contig of 2100 bp in length
* 12461 12560: gap of unknown length
* 12561 14611: contig of 2051 bp in length
* 14612 14711: gap of unknown length
* 14712 16381: contig of 1670 bp in length
* 16382 16481: gap of unknown length
* 16482 17968: contig of 1487 bp in length
* 17969 18068: gap of unknown length
* 18069 20434: contig of 2366 bp in length
```

```
* 20435 20534: gap of unknown length
* 20535 23515: contig of 2981 bp in length
* 23516 23615: gap of unknown length
* 23616 27563: contig of 3948 bp in length
* 27564 27663: gap of unknown length
* 27664 30987: contig of 3324 bp in length
* 30988 31087: gap of unknown length
* 31088 36065: contig of 4978 bp in length
* 36066 36165: gap of unknown length
* 36166 40978: contig of 4813 bp in length
* 40979 41078: gap of unknown length
* 41079 45663: contig of 4585 bp in length
* 45664 45763: gap of unknown length
* 45764 51745: contig of 5982 bp in length
* 51746 51845: gap of unknown length
* 51846 57359: contig of 5514 bp in length
* 57360 57459: gap of unknown length
* 57460 67881: contig of 10422 bp in length
* 67882 74132: contig of 6151 bp in length
* 74133 74232: gap of unknown length
* 74233 79795: contig of 5563 bp in length
* 79796 79895: gap of unknown length
* 79896 87511: contig of 7616 bp in length
* 87512 92791: contig of 5180 bp in length
* 92792 92891: gap of unknown length
* 92892 102794: contig of 9903 bp in length
* 102795 102894: gap of unknown length
* 102895 110866: contig of 7972 bp in length
* 110867 110966: gap of unknown length
* 110967 117571: contig of 6605 bp in length
* 117572 117671: gap of unknown length
* 117672 123738: contig of 6067 bp in length
* 123739 123838: gap of unknown length
* 123839 130583: contig of 6745 bp in length
* 130584 130683: gap of unknown length
* 130684 141544: contig of 10861 bp in length
* 141545 141644: gap of unknown length
* 141645 169109: contig of 27465 bp in length
* 169110 169209: gap of unknown length
* 169210 193446: contig of 24237 bp in length.
FEATURES
Location/Qualifiers
source
1..193446
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="RP11-360I2"
/clone_lib="RPCI human BAC library 11"
BASE COUNT 57571 a 35252 c 37061 g 60360 t 3202 others
ORIGIN
Query Match 40.9%; Score 62.6; DB 2; Length 193446;
Best Local Similarity 83.5%; Pred. No. 2.7e-08;
Matches 71; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
QY 52 AAGAACCCCGGCAAGCAGCTACAGACCCCGTGAGACTACTCTATTAATTAAT 111
DB 87679 AAAAAACCAAGCAAGCAAGTTGCAGACCCCGTGAGACTACTCTATTAATTAAT 111
QY 112 AAGACCTTGCAGAAAGGAAGCAC 136
DB 87739 AGAGATCTTGCACCAAGAGAGCTC 87763
```

Search completed: August 2, 2002, 20:30:20  
Job time: 25061 sec

Mon Aug 5 11:51:51 2002

us-08-973-363-3.rge

Page 9



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 2, 2002, 20:47:17 ; Search time 1205.34 Seconds  
(without alignments) 217.937 Million cell updates

Title: US-08-973-363-3  
 perfect score: 153  
 Sequence: 1 ATTTTACCTGATGATCCAGA.....CACAAAGCTTGCTGGTGA 153

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters:	3472872
--	---------

```

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                  Maximum Match 10%
                  Listing first 45

```

Database : N\_Genesed\_032802.\*

1:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1980.DAT.*
2:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1981.DAT.*
3:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1982.DAT.*
4:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1983.DAT.*
5:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1984.DAT.*
6:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1985.DAT.*
7:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1986.DAT.*
8:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1987.DAT.*
9:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1988.DAT.*
10:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1989.DAT.*
11:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1990.DAT.*
12:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1991.DAT.*
13:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1992.DAT.*
14:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1993.DAT.*
15:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1994.DAT.*
16:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1995.DAT.*
17:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1996.DAT.*
18:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1997.DAT.*
19:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT.*
20:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT.*
21:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT.*
22:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT.*
23:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT.*
24:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Query %			Length	DB	ID	Description
	Score	Match					
1	153	100.0	153	18	AAT42757	Chick CHD-1A gene f	
2	140.2	91.6	153	18	AAT42758	Chick CHD-W gene f	
3	132.2	86.4	153	18	AAT42759	Great tit CHD-W gene	
4	103	67.3	6608	18	AAT42751	Chicken CHD-1A gene	
5	102.4	66.9	153	18	AAT42756	Mouse CHD-1 gene (	
6	92.4	60.4	1316	18	AAT42754	Chicken CHD-W gene	
7	42.8	28.0	1311	19	AAV59280	Altered telomere r	
8	38.2	25.0	421	22	AAK88882	Human digestive sy	
9	38.2	25.0	421	22	AAI57603	Human colorectal c	

10	36.6	23.9	6240	23	ABL064442
11	36.6	23.9	9933	23	ABL064442
12	32.4	21.2	1269	23	AA351571
13	31.8	20.8	5068	22	AAF32508
14	31.2	20.4	1290	24	AA151676
15	31	20.3	3855	23	ABL13371
16	31	20.3	8781	23	ABL13370
17	30.6	20.0	6694	24	ABL33665
18	30.4	19.9	2004	17	AA133952
19	30.4	19.9	2004	18	AAV73997
20	30.4	19.9	2004	19	AAV16170
21	30.4	19.9	2010	15	AA074683
22	30.4	19.9	2489	20	AA231960
23	30.4	19.9	2576	17	AA133945
24	30.4	19.9	2576	18	AAV74006
25	30.4	19.9	2576	19	AAV16181
26	30.4	19.9	2655	15	AA074682
27	30.4	19.9	2655	17	AA133951
28	30.4	19.9	2655	17	AA133951
29	30.4	19.9	2655	18	AAV73996
30	30.4	19.9	2655	18	AAV74001
31	30.4	19.9	2655	19	AAV16169
32	30.4	19.9	2655	19	AAV16174
33	30.4	19.9	4031	17	AA133947
34	30.4	19.9	4031	18	AAV74010
35	30.4	19.9	4031	19	AAV16189
36	30.2	19.7	558	21	AA094041
37	30.2	19.7	927	18	AA066501
38	30.2	19.7	1294	18	AA133864
39	30.2	19.7	57296	22	AAK78170
40	30.2	19.7	57296	22	AAK78847
41	30.2	19.7	57296	22	AAK79364
42	30.2	19.7	57296	22	AAK86799
43	30.2	19.6	230	18	AAV73848
44	30.2	19.6	850	14	AA050992
45	30.2	19.6	2837	23	ABL11216

## ALIGNMENTS

RESULT 1  
AAT42757  
ID AAT42757 standard: DNA: 153 BP.

XX  
AC AAT42757:

XX  
DT 12-MAR-1997 (first entry)

XX  
DE Chick CHD-1A gene fragment.

XX Bird; sex determination; chromo

KW CHD-1A; CHD-W; W chromosome; SS  
XX

US  
XX  
Gallus sp.

FT	key	Location/Qualit
FT	misc_difference	52..81

```

FT      /"lay= d
FT      /note= "bases 5

```

FT and acid se

XX  
PN  
W09639505-A1.

PD 12-DEC-1996.

05-JUN-1996; 96WO-GB01341.

PR 06-JUN-1995; 95GB-0011439.

PA (ISIS-) ISIS INNOVATION LTD.

PI Griffiths R, Tiwari B;  
 XX WPI; 1997-043127/04.  
 DR P-PSDB; AAW08147.  
 XX Avian chromodomain-helicase-DNA binding genes determine sex in  
 PT birds - used for sex determin. and to control sex of progeny  
 XX  
 PS Claim 8; Fig 3; 76pp; English.  
 XX  
 CC Bases 3855-3977 (AAW42756) of the mouse CHD-1 gene show homology  
 CC to portions of the chicken CHD-1A (A = Avian) gene (AAW42757),  
 CC chicken CHD-W (W refers to the W chromosome) gene (AAW42758) and  
 CC and the great tit CHD-W gene (see also AAT42759). Translated amino  
 CC acid sequences of this region are provided in AAW08146-49. The  
 CC CHD-1A (see also AAT42751) and CHD-W (see also AAT42754-55) genes  
 CC determine sex in birds and can be used to identify the sex of an  
 CC embryo, foetus etc. and to manipulate the sex of progeny.  
 XX  
 SQ Sequence 153 BP; 58 A; 40 C; 31 G; 24 T; 0 other;

Query Match 100.08; Score 153; DB 18; Length 153;  
 Best Local Similarity 100.08; Pred. No. 1.6e-37;  
 Matches 153; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATTTTACCTGATGATCCAGACAAAGAACCCAGGCAAGCAGCTACAGACCAAGAAACCC 60  
 DB 1 attttacctgatgatccagacagaagaacccccaggaagcagctacagaccagaagaaccc 60  
 QY 61 CAGGCAAGCAGCTACAGACCCGTCGAGACTACCTCTAAATTAATTAATTAAGACCTT 120  
 DB 61 caggcaagcagctacagaccgcgtgcagactaccttaataataactgaataaagacctt 120  
 QY 121 GCAAGAAAGGAGCACAAGGCTTCTGCTGCA 153  
 DB 121 gcaagaaaggagcacaagggcttctgctgtgca 153

RESULT 2  
 AAT42758  
 ID AAT42758 standard; DNA; 153 BP.  
 XX  
 AC AAT42758;  
 XX  
 DT 12-MAR-1997 (first entry)  
 XX  
 DE Chick CHD-W gene fragment.  
 XX  
 KW Bird; sex determination; chromodomain-Helicase-DNA binding 1;  
 KW CHD-1A; CHD-W; W chromosome; ss.  
 XX  
 OS Gallus sp.  
 XX  
 FH Key Location/Qualifiers  
 FT misc\_difference 52..81  
 FT /\*tag= a  
 FT /note= "bases 52-81 are a repeat of bases 22-51  
 FT and are ignored in the translated amino  
 FT acid sequence given in Fig 3"  
 XX  
 PN WO9639505-A1.  
 XX  
 PD 12-DEC-1996.  
 XX  
 PF 05-JUN-1996; 96WO-GB01341.  
 XX  
 PR 06-JUN-1995; 95GB-0011439.  
 XX  
 PA (ISIS-) ISIS INNOVATION LTD.  
 XX  
 PI Griffiths R, Tiwari B;  
 XX  
 DR WPI; 1997-043127/04.  
 DR P-PSDB; AAW08149.

DR WPI; 1997-043127/04.  
 DR P-PSDB; AAW08148.  
 XX  
 PT Avian chromodomain-helicase-DNA binding genes determine sex in  
 PT birds - used for sex determin. and to control sex of progeny  
 XX  
 PS Claim 8; Fig 3; 76pp; English.  
 XX  
 CC Bases 3855-3977 (AAW42756) of the mouse CHD-1 gene show homology  
 CC to portions of the chicken CHD-1A (A = Avian) gene (AAW42757),  
 CC chicken CHD-W (W refers to the W chromosome) gene (AAW42758) and  
 CC and the great tit CHD-W gene (AAW42759). Translated amino acid  
 CC sequences of this region are provided in AAW08146-49. The CHD-1A  
 CC (see also AAT42751) and CHD-W (see also AAT42754-55) genes determine  
 CC sex in birds and can be used to identify the sex of an embryo,  
 CC foetus etc. and to manipulate the sex of progeny.  
 XX  
 SQ Sequence 153 BP; 56 A; 36 C; 31 G; 30 T; 0 other;

Query Match 91.6%; Score 140.2; DB 18; Length 153;  
 Best Local Similarity 94.8%; Pred. No. 1.3e-33;  
 Matches 145; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 ATTTTACCTGATGATCCAGACAAAGAACCCAGGCAAGCAGCTACAGACCAAGAAACCC 60  
 DB 1 attttacctgatgatccagacagaagaacccccaggaagcagctacagaccagaagaaccc 60  
 QY 61 CAGGCAAGCAGCTACAGACCCGTCGAGACTACCTCTAAATTAATTAATTAAGACCTT 120  
 DB 61 caggcaagcagctacagaccgcgtgcagactaccttaataataactgaataaagacctt 120  
 QY 121 GCAAGAAAGGAGCACAAGGCTTCTGCTGCA 153  
 DB 121 gcaagaaaggagcacaagggcttctgctgtgca 153

RESULT 3  
 AAT42759  
 ID AAT42759 standard; DNA; 153 BP.  
 XX  
 AC AAT42759;  
 XX  
 DT 12-MAR-1997 (first entry)  
 XX  
 DE Great tit CHD-W gene fragment.  
 XX  
 KW Bird; sex determination; chromodomain-Helicase-DNA binding 1;  
 KW CHD-1A; CHD-W; W chromosome; ss.  
 XX  
 OS Parus major.  
 XX  
 FH Key Location/Qualifiers  
 FT misc\_difference 52..81  
 FT /\*tag= a  
 FT /note= "bases 52-81 are a repeat of bases 22-51  
 FT and are ignored in the translated amino  
 FT acid sequence given in Fig 3"  
 XX  
 PN WO9639505-A1.  
 XX  
 PD 12-DEC-1996.  
 XX  
 PF 05-JUN-1996; 96WO-GB01341.  
 XX  
 PR 06-JUN-1995; 95GB-0011439.  
 XX  
 PA (ISIS-) ISIS INNOVATION LTD.  
 XX  
 PI Griffiths R, Tiwari B;  
 XX  
 DR WPI; 1997-043127/04.  
 DR P-PSDB; AAW08149.



XX	Avian chromodomain-helicase-DNA binding genes determine sex in birds - used for sex determ.
PT	and to control sex of progeny
PT	
XX	Claim 8; Fig 3; 76pp; English.
XX	Bases 3855-3977 (AAT42756) of the mouse CHD-1 gene show homology to portions of the chicken CHD-1A (A = Avian) gene (AAT42757), chicken CHD-W (W refers to the W chromosome) gene (AAT42758) and the great tit CHD-W gene (AAT42759). Translated amino acid sequences of this region are provided in AA008146-49. The CHD-1A (see also AAT42751) and CHD-W (see also AAT42754-55) genes determine sex in birds and can be used to identify the sex of an embryo, foetus etc. and to manipulate the sex of progeny.
XX	Sequence 153 BP; 60 A; 33 C; 31 G; 29 T; 0 other;
SQ	
Query Match	86.4%; Score 132.2; DB 18; Length 153;
Best Local Similarity	91.5%; Pred. No. 3.6e-31;
Matches 140;	Conservative 0; Mismatches 13; Indels 0; Gaps
QY	1 ATTTTACCTGTGATGCAGACAAGAACCACCGCAAGCAGCTACGACCAACGAACCC 60
Ddb	1 attttacctgagtgcagataagaacaccaggccaagcagtgtcgacacaagaacca 60
QY	61 CAGCAAAACGACTACAGACCCGTGCAGACTACCTTCATTAAATTACTGAATAAGACCTT 120
Ddb	61 caggcaaaacagttgcagaccctgcagattacctctattaattactgaataaagacctt 120
QY	121 GCAGAAGGAAGAACACAAAGCGTTGCTGGTGCA 153
Ddb	121 qcaagaagaagatgtcgaadgaacttactgtaca 153

RESULT	4
AA142751	AA142751 standard; cDNA; 6508 BP.
XX	
AC	AA142751;
XX	
DT	12-MAR-1997 (first entry)
XX	
DE	Chicken CHD-1a gene.
XX	
KW	Bird; sex determination; chromodomain-Helicase-DNA binding 1 Avian;
XX	
KW	CHD-W; chromodomain-Helicase-DNA binding on the W chromosome; ss.
XX	
OS	Gallus sp.
XX	
Key	Location/Qualifiers
FFH	228..5390
FT	/*tag= a
XX	
FT	
XX	
PN	W09639505-A1.
XX	
PD	12-DEC-1996.
XX	
PF	05-JUN-1996; 96WO-GB01341.
XX	
PR	06-JUN-1995; 95GB-0011439.
XX	
PA	(ISIS-) ISIS INNOVATION LTD.
XX	
PI	Griffiths R, Tiwari B;
XX	
DR	WPI; 1997-043127/04.
XX	
PPT	Avian chromodomain-helicase-DNA binding genes determine sex in
PPT	birds - used for sex determ. and to control sex of progeny
XX	
PS	Claim 1; Fig 5; 76pp; English.

CC The chicken CHD-W gene (AA742754) acting alone or in conjunction with  
CC the closely related CHD-1A gene (AA742751) is suggested to initiate  
CC female development in birds. The sequence of CHD-1A was deduced  
CC from 3 clones isolated from a stage 10-12 chicken embryo cDNA  
CC library using a great tit CHD-W sequence (see also AA742755) as probe.  
CC The CHD-1A (A = Avian) gene shows close identity to the mouse CHD-1  
CC gene (see also AA742756-57). It is located on an autosome or 2  
CC chromosome. Probes based on CHD-W and CHD-1A give a W chromosome-  
CC specific signal on hybridisation to genomic DNA of a non-rare  
CC bird and can be used for sex determin. of a bird. CHD-1A nucleic  
CC acids can also be used to control the sex of the progeny of a bird.  
XX  
SQ Sequence 6608 BP; 2289 A; 1207 C; 1459 G; 1653 T; 0 other;

Query Match 67.3%; Score 103; DB 18; Length 6608;  
Best Local Similarity 100.0%; Pred. No. 9.9e-22;  
Matches 103; Conservative 0; Mismatches 0; Indels 0; Gaps

QY 51 CAAGAACCAGGCAAGCAGCTACAGACCCGTGCAGACTACCTCATTAATTAAGTCAA 110  
DB 4100 caagaaccagcagaagcagctacagaccgcgtagactaccattcattactgaa 4159  
|||||  
QY 111 TAAAGACCTTCGACGAAGGAGCACAAGGCTTGTCTGGTGCA 153  
DB 4160 taaagaccttcgaagaaggaagcacaagaagcttgctggtgca 4202  
|||||

RESULT 5  
ID AA742756 standard; DNA; 153 BP.  
XX  
AC AA742756;  
XX  
DT 12-MAR-1997 (first entry)  
XX  
DE Mouse CHD-1 gene (bases 3855-977).  
XX  
KW Bird; sex determination; chromodomain-Helicase-DNA binding 1;  
KW CHD-1; CHD-W; W chromosome; ss.  
XX  
OS Mus sp.  
XX  
FH Key Location/Qualifiers  
FT misc\_difference 52..81  
FT /\*tag= a  
FT /\*note= "bases 52-81 are a repeat of bases 22-51  
FT and are ignored in the translated amino  
FT acid sequence given in Fig 3".  
XX  
PN W09639505-A1.  
XX  
PD 12-DEC-1996.  
XX  
PP 05-JUN-1996; 96WO-GB01341.  
XX  
PR 06-JUN-1995; 95GB-0011439.  
XX  
PA (ISIS-) ISIS INNOVATION LTD.  
XX  
PI Griffiths R, Tiwari B;  
XX  
XX WPI; 1997-043127/04.  
XX P-PSDB; AAW08146.  
XX  
XX Avian chromodomain-helicase-DNA binding genes determine sex in  
XX birds - used for sex determin. and to control sex of progeny  
XX  
XX Claim 8; Fig 3; 76pp; English.  
XX  
XX Bases 3855-3977 (AA742756) of the mouse CHD-1 gene show homology  
XX to portions of the chicken CHD-1A (A = Avian) gene (see also  
XX and AA742757). chicken CHD-W (W refers to the W chromosome) gene

CC and AAT42757), chicken CHD-W (W refers to the W chromosome) gene

CC (see also AAT42758) and the great tit CHD-W gene (see also AAT42759).  
CC Translated amino acid sequences of this region are provided in  
CC AAW08146-49. The CHD-1A (see also AAT42751) and CHD-W (see also  
CC AAT42754-55) genes determine sex in birds and can be used to identify  
CC the sex of an embryo, foetus etc. and to manipulate the sex of  
CC progeny.

XX Sequence 153 BP; 61 A; 37 C; 27 G; 28 T; 0 other;

Query Match 66.9%; Score 102.4; DB 18; Length 153;  
Best Local Similarity 79.6%; Pred. No. 4.8e-22;  
Matches 121; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 1 ATTTTACCTGATGATCCAGACAGAACCCAGGCAAGCAGCTACAGACCAAGAACCC 60  
DB 1 attcttccagatgattcctgataaaaccacagcaagctacagaccacaaaccca 60  
QY 61 CAGGCAAGCAGCTACAGACCCGCTGCAGACTACCTCATTTAAATTACTGAATAAGACCTT 120  
DB 61 caagcaaaacagttacagaccctgacagactacctcatcaactacttagcagagatctt 120  
QY 121 GCAAGAAAGGAGCAACAAAGGCTTGCTGGTGC 152  
DB 121 gcaaaagagaggtcagagactttgtggtgc 152

RESULT 6  
AAT42754  
ID AAT42754 standard; cDNA; 1316 BP.  
XX  
AC AAT42754;  
XX  
DT 12-MAR-1997 (first entry)  
XX  
DE Chicken CHD-W gene (partial sequence).  
XX  
KW Bird; sex determination; chromodomain-Helicase-DNA binding 1 Avian;  
KW CHD-W; chromodomain-Helicase-DNA binding on the W chromosome; ss.  
OS Gallus sp.  
XX  
PN WO9639505-Al.  
XX  
PD 12-DEC-1996.  
XX  
PF 05-JUN-1996; 96WO-GB01341.  
XX  
PR 06-JUN-1995; 95GB-0011439.  
XX  
PA (ISIS-) ISIS INNOVATION LTD.  
XX  
PI Griffiths R, Tiwari B;  
XX  
DR WPI; 1997-043127/04.  
XX  
PT Avian chromodomain-helicase-DNA binding genes determine sex in  
PT birds - used for sex determin. and to control sex of progeny  
XX  
PS Claim 1; Fig 8; 76pp; English.

XX The chicken CHD-W gene (AAT42754) acting alone or in conjunction with  
XX the closely related CHD-1A gene (AAT42751) is suggested to initiate  
XX female development in birds. The sequence of CHD-W was deduced  
XX from 2 clones isolated from a 10-day chick embryo library using  
XX a fragment of the CHD-1A gene as a probe. The CHD-W gene is  
XX located on the W chromosome. Probes based on CHD-W and CHD-1A give  
XX a W chromosome-specific signal on hybridisation to genomic DNA of a  
XX non-rare bird and can be used for sex determin. of a bird. CHD-W  
XX nucleic acids can also be used to control the sex of progeny of a  
XX bird.

XX Sequence 1316 BP; 492 A; 208 C; 306 G; 304 T; 6 other;

Query Match 60.4%; Score 92.4; DB 18; Length 1316;  
Best Local Similarity 94.1%; Pred. No. 1.1e-18;  
Matches 96; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 52 AAGAAACCCAGCAAGCAGCTACAGACCCGCTGCAGACTACCTCATTAATTAATTAAT 111  
DB 1159 aagaaacccagcagcagctacagactctgagattacaccttaataataactgaat 1218  
QY 112 AAGACCTTGCAGAAAGAACACAAAGGCTTGCTGGTGC 153  
DB 1219 aagaccttgcagaaaggacacagagacttgcgtgtgca 1260

RESULT 7  
AAV59280  
ID AAV59280 standard; cDNA; 1311 BP.  
XX  
AC AAV59280;  
XX  
DT 14-DEC-1998 (first entry)  
XX  
DE Altered telomere repeat binding factor 1 gene.  
XX  
KW ds; human; telomere repeat binding factor; A-TRF; dimerisation domain;  
KW telomere; ageing; ataxia telangiectasia; Down's syndrome; tumour; viral.  
XX  
OS Homo sapiens.  
XX  
FH Key  
FT CDS  
FT 1..1311  
FT /\*tag= a  
FT /product= "A-TRF"  
XX  
PN WO9836066-Al.  
XX  
PD 20-AUG-1998.  
XX  
PF 13-FEB-1998; 98WO-US02765.  
XX  
PR 04-FEB-1998; 98US-0018628.  
XX  
PR 13-FEB-1997; 97US-0800264.  
XX  
PA (UYRQ ) UNIV ROCKEFELLER.  
XX  
PI Bianchi A, De Lange T, Van Steensel B;  
XX  
DR WPI; 1998-480769/41.  
XX  
DR P-PSDB; AAV59280.  
XX  
PT Nucleic acid encoding altered telomere repeat binding protein and  
PT related vectors - transformants, hetero-dimers and antibodies, used  
PT to inhibit shortening of telomerases caused by ageing or disease,  
PT also used to extend life of cells in culture  
XX  
PS Claim 14; Page 110-111; 163pp; English.

XX The altered vertebrate telomere repeat binding protein (A-TRF) has a  
XX telomere repeat binding factor (TRF) dimerisation domain, and forms a  
XX hetero-dimer with TRF, preventing it from binding to the specified repeat  
XX sequence. A-TRF, optionally expressed by gene therapy, is used to  
XX inhibit shortening of telomeres associated with ageing (for cosmetic  
XX purposes) and disease, e.g. ataxia telangiectasia, Down's syndrome,  
XX atrophy of the skin, age-related macular degeneration, atherosclerosis,  
XX tumours and viral (including human immune deficiency virus) infection.  
XX Cells expressing A-TRF also have an increased life span in vitro, e.g.  
XX for expression of recombinant proteins or where intended for subsequent  
XX transplant or for testing, eliminating the need for transformation.

XX Sequence 1311 BP; 441 A; 257 C; 316 G; 297 T; 0 other;

Query Match 28.0%; Score 42.8; DB 19; Length 1311;  
 Best Local Similarity 87.0%; Pred. No. 0.0016;  
 Matches 47; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
 Oy 52 AAGAACCCAGGCAAGCAGCTACAGACCCGTCGAGACTACCTCATTAATTA 105  
 || |||| || |||| || |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 Db 1258 aaaaacacaaagcaaacaggtacagaccgtgcagactacccatcaacta 1311

RESULT 8  
 AAK88882  
 ID AAK8882 standard; cDNA; 421 BP.  
 AC AAK8882;  
 DT 05-NOV-2001 (first entry)  
 XX Human digestive system antigen coding sequence SEQ ID NO: 1198.  
 XX Human; digestive system antigen; gene therapy; cancer; appendicitis;  
 KW ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;  
 KW digestive system disorder; Meckel's diverticulum; ss.  
 XX Homo sapiens.  
 XX WO200155314-A2.  
 XX 02-AUG-2001.  
 XX 17-JAN-2001; 2001WO-US01324.

PR 31-JAN-2000; 2000US-0179065.  
 PR 04-FEB-2000; 2000US-0180628.  
 PR 24-FEB-2000; 2000US-0184664.  
 PR 02-MAR-2000; 2000US-0186350.  
 PR 16-MAR-2000; 2000US-0189874.  
 PR 17-MAR-2000; 2000US-0190070.  
 PR 18-APR-2000; 2000US-0198123.  
 PR 13-MAY-2000; 2000US-0205515.  
 PR 07-JUN-2000; 2000US-0209467.  
 PR 28-JUN-2000; 2000US-0214886.  
 PR 30-JUN-2000; 2000US-0215135.  
 PR 07-JUL-2000; 2000US-0216647.  
 PR 07-JUL-2000; 2000US-0216880.  
 PR 11-JUL-2000; 2000US-0217487.  
 PR 14-JUL-2000; 2000US-0217496.  
 PR 26-JUL-2000; 2000US-0218290.  
 PR 26-JUL-2000; 2000US-0220963.  
 PR 14-AUG-2000; 2000US-0220964.  
 PR 14-AUG-2000; 2000US-0224518.  
 PR 14-AUG-2000; 2000US-0224519.  
 PR 14-AUG-2000; 2000US-0225213.  
 PR 14-AUG-2000; 2000US-0225214.  
 PR 14-AUG-2000; 2000US-0225266.  
 PR 14-AUG-2000; 2000US-0225267.  
 PR 14-AUG-2000; 2000US-0225268.  
 PR 14-AUG-2000; 2000US-0225270.  
 PR 14-AUG-2000; 2000US-0225447.  
 PR 14-AUG-2000; 2000US-0225757.  
 PR 14-AUG-2000; 2000US-0225758.  
 PR 18-AUG-2000; 2000US-0225759.  
 PR 18-AUG-2000; 2000US-0226279.  
 PR 22-AUG-2000; 2000US-0226681.  
 PR 22-AUG-2000; 2000US-0226688.  
 PR 23-AUG-2000; 2000US-0227182.  
 PR 30-AUG-2000; 2000US-0227009.  
 PR 01-SEP-2000; 2000US-0228924.  
 PR 01-SEP-2000; 2000US-0229287.  
 PR 01-SEP-2000; 2000US-0229343.  
 PR 01-SEP-2000; 2000US-0229344.  
 PR 01-SEP-2000; 2000US-0229345.  
 PR 05-SEP-2000; 2000US-0229509.  
 PR 05-SEP-2000; 2000US-0229513.  
 PR 06-SEP-2000; 2000US-0230437.  
 PR 06-SEP-2000; 2000US-0230438.  
 PR 08-SEP-2000; 2000US-0231242.  
 PR 08-SEP-2000; 2000US-0231243.  
 PR 08-SEP-2000; 2000US-0231244.  
 PR 08-SEP-2000; 2000US-0231245.  
 PR 08-SEP-2000; 2000US-0231413.  
 PR 08-SEP-2000; 2000US-0231414.  
 PR 08-SEP-2000; 2000US-0232080.  
 PR 08-SEP-2000; 2000US-0232081.  
 PR 12-SEP-2000; 2000US-0231968.  
 PR 14-SEP-2000; 2000US-0232397.  
 PR 14-SEP-2000; 2000US-0232398.  
 PR 14-SEP-2000; 2000US-0232399.  
 PR 14-SEP-2000; 2000US-0232400.  
 PR 14-SEP-2000; 2000US-0232401.  
 PR 14-SEP-2000; 2000US-0233063.  
 PR 14-SEP-2000; 2000US-0233064.  
 PR 14-SEP-2000; 2000US-0233065.  
 PR 21-SEP-2000; 2000US-0234223.  
 PR 21-SEP-2000; 2000US-0234274.  
 PR 25-SEP-2000; 2000US-0234997.  
 PR 25-SEP-2000; 2000US-0234998.  
 PR 26-SEP-2000; 2000US-0235484.  
 PR 27-SEP-2000; 2000US-0235834.  
 PR 27-SEP-2000; 2000US-0235836.  
 PR 29-SEP-2000; 2000US-0236327.  
 PR 29-SEP-2000; 2000US-0236367.  
 PR 29-SEP-2000; 2000US-0236368.  
 PR 29-SEP-2000; 2000US-0236369.  
 PR 29-SEP-2000; 2000US-0236370.  
 PR 02-OCT-2000; 2000US-0236802.  
 PR 02-OCT-2000; 2000US-0237037.  
 PR 02-OCT-2000; 2000US-0237038.  
 PR 02-OCT-2000; 2000US-0237039.  
 PR 02-OCT-2000; 2000US-0237040.  
 PR 13-OCT-2000; 2000US-0239935.  
 PR 13-OCT-2000; 2000US-0239937.  
 PR 20-OCT-2000; 2000US-0240960.  
 PR 20-OCT-2000; 2000US-0241121.  
 PR 20-OCT-2000; 2000US-0241785.  
 PR 20-OCT-2000; 2000US-0241786.  
 PR 20-OCT-2000; 2000US-0241787.  
 PR 20-OCT-2000; 2000US-0241808.  
 PR 20-OCT-2000; 2000US-0241809.  
 PR 20-OCT-2000; 2000US-0241826.  
 PR 01-NOV-2000; 2000US-0244617.  
 PR 08-NOV-2000; 2000US-0246474.  
 PR 08-NOV-2000; 2000US-0246475.  
 PR 08-NOV-2000; 2000US-0246476.  
 PR 08-NOV-2000; 2000US-0246477.  
 PR 08-NOV-2000; 2000US-0246478.  
 PR 08-NOV-2000; 2000US-0246523.  
 PR 08-NOV-2000; 2000US-0246524.  
 PR 08-NOV-2000; 2000US-0246525.  
 PR 08-NOV-2000; 2000US-0246526.  
 PR 08-NOV-2000; 2000US-0246527.  
 PR 08-NOV-2000; 2000US-0246528.  
 PR 08-NOV-2000; 2000US-0246532.  
 PR 08-NOV-2000; 2000US-0246609.  
 PR 08-NOV-2000; 2000US-0246610.  
 PR 08-NOV-2000; 2000US-0246611.  
 PR 08-NOV-2000; 2000US-0246613.  
 PR 17-NOV-2000; 2000US-0249207.  
 PR 17-NOV-2000; 2000US-0249208.  
 PR 17-NOV-2000; 2000US-0249209.  
 PR 17-NOV-2000; 2000US-0249210.  
 PR 17-NOV-2000; 2000US-0249211.  
 PR 17-NOV-2000; 2000US-0249212.  
 PR 17-NOV-2000; 2000US-0249213.  
 PR 17-NOV-2000; 2000US-0249214.  
 PR 17-NOV-2000; 2000US-0249215.  
 PR 17-NOV-2000; 2000US-0249216.





XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins  
CC (ABB57737-ABB72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 6240 BP; 1675 A; 1682 C; 1760 G; 1123 T; 0 other;

Query Match 23.9%; Score 36.6; DB 23; Length 6240;  
Best Local Similarity 71.6%; Pred. No. 0.21;  
Matches 48; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 55 AAACCCAGGCAAGCAGCTACAGACCGTGCAGACTACCTCATTAATAATTAATGAATAAA 114  
II IIIIIII IIIIIII IIIII IIIII IIIII III IIIII III IIIII  
DB 4205 aagccccagggcagcagctgcagcgcgtccgagtgctcagatcatcaagaag 4264

QY 115 GACCTTG 121  
II III  
DB 4265 aacgtgg 4271

RESULT 11  
ID ABL06442 standard; cDNA; 9933 BP.  
XX  
AC ABL06442;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 13808.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical; gene; ss.  
XX  
OS Drosophila melanogaster.  
XX  
PN WO200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US09231.  
XX  
PR 23-MAR-2000; 2000US-191637P.  
PR 11-JUL-2000; 2000US-0614150.  
XX  
PA (PEKE ) PE CORP NY.  
XX  
PI Venter JC, Adams M, Li PWD, Myers EW;  
XX  
DR WPI; 2001-656860/75.  
DR P-PSDB; ABB62339.  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX  
PS Claim 1; SEQ ID NO 13808; 21pp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins

CC (ABB57737-ABB72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 9933 BP; 2711 A; 2468 C; 2538 G; 2216 T; 0 other;

Query Match 23.9%; Score 36.6; DB 23; Length 9933;  
Best Local Similarity 71.6%; Pred. No. 0.24;  
Matches 48; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 55 AAACCCAGGCAAGCAGCTACAGACCGTGCAGACTACCTCATTAATAATTAATGAATAAA 114  
II IIIIIII IIIIIII IIIII IIIII IIIII III IIIII III IIIII  
DB 6745 aagccccagggcagcagctgcagcgcgtccgagtgctcagatcatcaagaag 6804

QY 115 GACCTTG 121  
II III  
DB 6805 aacgtgg 6811

RESULT 12  
AAS51571/C  
ID AAS51571 standard; DNA; 1269 BP.  
XX  
AC AAS51571;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE Pseudomonas aeruginosa DNA for cellular proliferation protein #156.  
XX  
KW Antisense; ds; prokaryotic cellular proliferation gene;  
KW antibiotic; antibacterial; drug design.  
XX  
OS Pseudomonas aeruginosa.  
XX  
PN WO200170955-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 21-MAR-2001; 2001WO-US09180.  
XX  
PR 21-MAR-2000; 2000US-191078P.  
PR 23-MAY-2000; 2000US-206848P.  
PR 26-MAY-2000; 2000US-207727P.  
PR 23-OCT-2000; 2000US-242578P.  
PR 27-NOV-2000; 2000US-253625P.  
PR 22-DEC-2000; 2000US-257931P.  
PR 16-FEB-2001; 2001US-269308P.  
XX  
PA (ELIT-) ELITRA PHARM INC.  
XX  
PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;  
PI Yamamoto RT, Xu HH;  
XX  
DR WPI; 2001-611495/70.  
DR P-PSDB; AAU33712.  
XX  
PT New polynucleotides for the identification and development of  
PT antibiotics, comprise sequences of antisense nucleic acids -  
XX  
PS Claim 27; Seq ID NO 4153; 511pp; English.  
XX  
CC The invention relates to antisense inhibitors of genes essential to  
CC prokaryotic cellular proliferation, their use in identifying the  
CC genes, their use in the discovery of novel antibiotics, the essential  
CC genes themselves and the encoded proteins. The prokaryotes used are  
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella  
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The  
CC invention is also useful for the identification of potential new targets  
CC for antibiotic development. The antisense nucleic acids can also be used  
CC to identify proteins used in proliferation, to express these proteins,  
CC and to obtain antibodies capable of binding to the expressed proteins.

CC The proteins can be used to screen compounds in rational drug discovery  
CC programmes. The antisense nucleic acid sequence is also useful to screen  
CC for homologous nucleic acids which are required for cell proliferation in  
CC a wide variety of organisms. The present sequence encodes an  
CC essential prokaryotic cellular proliferation protein.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ

Sequence 1269 BP; 135 A; 436 C; 436 G; 262 T; 0 other;

Query Match 21.2%; Score 32.4; DB 23; Length 1269;  
Best Local Similarity 64.9%; Pred. No. 2.5; Mismatches 0; Gaps 0;  
Matches 48; Conservative 0; Mismatches 26; Indels 0;  
OY 17 CAGACAGAAACCCAGGCAAGAGCTACAGACCAAGAAACCCAGGCAAGCTAC 76  
DB 668 CCGGCGACAGCGCCAGGCGGCTGTGGGCGGAAAGCTCCAGGCGACATGCTTC 609  
OY 77 AGACCGGTGCAGAC 90  
DB 608 AGGCCCTTGCCGAC 595

RESULT 13  
AAF32508  
ID AAF32508 standard; cDNA; 5068 BP.  
AC AAF32508;  
XX  
DT 19-APR-2001 (first entry)  
DE Human male enhanced antigen-2 (MEA-2) nucleotide sequence SEQ ID NO:1.  
XX  
KW Human; male enhanced antigen-2; MEA-2; identification; spermatogenesis;  
KW spermatogenesis disease; chromosome marker; pancreatic cancer; ss.  
XX  
OS Homo sapiens.  
XX  
PN JP2000316580-A.  
PD 21-NOV-2000.  
XX  
PF 30-APR-1999; 99JP-0125196.  
XX  
PR 30-APR-1999; 99JP-0125196.  
XX  
PA (ITOH-) ITO HAM KK.  
XX  
XX WPI; 2001-128256/14.  
DR P-PSDB; AAB69070.  
XX

A new protein, human male-enhanced antigen-2, useful for detecting  
spermatogenesis diseases  
Claim 2; Page 10-11; 21pp; Japanese.

The present sequence encodes the human male enhanced antigen-2 (MEA-2).  
The present invention also described an antibody specific for the  
MEA-2 protein. The antibody can be used for the identification of a  
gene causing diseases related to spermatogenesis. The MEA-2 nucleotide  
sequence is useful as a chromosome marker, and in the detection of  
pancreatic cancer.

Sequence 5068 BP; 1278 A; 1311 C; 1607 G; 872 T; 0 other;

Query Match 20.8%; Score 31.8; DB 22; Length 5068;  
Best Local Similarity 71.2%; Pred. No. 5.8;  
Matches 42; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

OY 37 AACGACGTACAGACCAAGAAACCCAGGCAAGCTACAGACCCGTCGAGACTACCT 95  
DB 3942 aaggagctgcaggcccaagggaacacactggtgcagaagctgcaggccgagccgagacct 4000

## RESULT 14

AAS15676  
ID AAS15676 standard; cDNA; 1290 BP.

AC AAS15676;

XX 29-JAN-2002 (first entry)

XX Human cDNA encoding prostate-specific membrane antibody protein 12.

XX Human; ss; prostate-specific membrane antibody protein 12; cytostatic;  
KW virucidal; immunomodulatory; antiinflammatory; haemostatic; cancer;  
KW haemopathy; human immunodeficiency virus infection; HIV;  
KW immunological disease; inflammation; prostate carcinoma;  
KW benign prostate tumour; metabolic disease; folic acid deficiency;  
KW neuropsychosis; developmental disturbance.

OS Homo sapiens.

XX Key Location/Qualifiers

FT CDS 965..1288

FT FT /\*tag= a

FT FT /product= "Prostate-specific membrane antibody  
protein 12"

XX WO200174892-A1.

XX 11-OCT-2001.

XX 19-MAR-2001; 2001WO-CN00351.

XX 22-MAR-2000; 2000CN-0115032.

XX (BIOW-) BIOWINDOW GENE DEV INC SHANGHAI.

XX Mao Y, Xie Y;

XX WPI; 2002-025812/03.

DR P-PSDB; AAU10288.

XX Human prostate-specific membrane antibody protein 12 and encoded  
PT polynucleotide, applicable in diagnosis and treatment of malignant  
PT tumour, haemopathy, HIV infection, immunological diseases and various  
PT inflammations

XX Claim 6; Page 30; 35pp; Chinese.

XX The invention relates to an isolated polypeptide of human  
XX prostate-specific membrane antibody protein 12 its fragment, analogue,  
XX derivative and nucleic acid encoding it. The polypeptide, nucleic acid  
XX encoding it and an antibody raised against it are applicable in  
XX diagnosis and treatment of cancer, haemopathy, HIV (human  
XX immunodeficiency virus) infection, immunological diseases, various  
XX inflammations including, carcinoma of the prostate and benign tumour of  
XX the prostate, diseases due to metabolic defect of folic acid,  
XX neuropsychosis and developmental disturbance. The present sequence  
XX encodes the prostate-specific membrane antibody protein 12.

XX Sequence 1290 BP; 328 A; 300 C; 276 G; 386 T; 0 other;

Query Match 20.4%; Score 31.2; DB 24; Length 1290;  
Best Local Similarity 54.3%; Pred. No. 5.8;  
Matches 63; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

OY 2 TTTTACCTGATGATCCAGCAAGAAACCCAGGCAAGCTACAGACCAAGAAACCCC 61  
DB 766 ttattgctgtaggaacctgtgaaaggacctgcatctctggaacaaacacacacctta 825

Job time: 25931 sec

**Qy** 62 AGGAAAGCAGCTACAGACCCCGTCAGACTACCTCATTAATTACTGAATAAGAC 117  
||||| ||||| - ||| ||| | - ||| ||| | ||||| | - |||  
**Db** 826 tggcagtgcaggrcccccacccccctgaaaggcatcacactacttgaactgaatgaccac 881

RESULT 15

ABL13371  
ID ABL13371 standard; cDNA: 3885 BP.

AC ABL13371:

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster expressed polynucleotide SEO ID NO 34595.

KW Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical; gene; ss.

OS *Drosophila melanogaster*.

PN WO200171042-A2.

AA 27-SEP-2001

23-MAR-2001: 2001WO-US09231.

23-MAR-2000: 2000US-191637P.

FR XX  
11 JUL 2000; 200003-0014130.

FA (FERE ) FE CONF NI.  
XX  
XX

XX  
XX  
F I  
VENCER OC, ADAMS M, LI FWD, MYERS EW,

DR P-PSDB; ABB69268.

New isolated nucle

PT interactions -  
XX genes from *Staphylococcus aureus* and for elucidating cell signaling and cell-cell interactions  
PS Claim 1; SEQ ID NO 34595; 21pp + Sequence Listing; English.

CC The invention relates to an isolated nucleic acid detection

CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABU16176-ABL30511), expressed DNA  
CC sequences (ABU1840-ABU16175) and the encoded proteins  
CC (ABB57737-ABB72072)

CC  
CC  
CC  
CC

This patent did not form part of the printed and sequence data for this patent in electronic format directly from WIPO specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published pctsequences.

Sequence 3885 BP; 1065 A; 1008 C; 1182 G; 630 T; 0 other;

Query Match 20.3%; Score 31; DB 23; Length 3885;  
Best Local Similarity 64.8%; Pred. No. 9.4;  
Matches 46; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

QY 19 GACAAGAAACCCAGGCAAGCAGCTACAGACCAGAAACCCAGGCAAGCAGCTACAG 78

Db 1051 gataaccaagtctgcgtcaggagagctaaagaccagcatagcccaggccaagttccgacag 1110

Qy 79 ACCCGTGCAGA 89

Db 1111 gccattgccga 1121

Search completed: August 2, 2002, 20:47:20



Mon Aug 5 11:51:51 2002

us-08-973-363-3.rng

Page 11

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 2, 2002, 16:52:12 ; Search time 8940.62 seconds  
(without alignments)  
230.972 Million cell updates/sec

Title: US-08-973-363-3

Perfect score: 153

Sequence: 1 ATTTCACGTGATCCAG.....CACAAAGCGTGTGCTGCA 153

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 13736207 segs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST.\*

- 1: em\_estba.\*
- 2: em\_esthum.\*
- 3: em\_estin.\*
- 4: em\_estnu.\*
- 5: em\_estov.\*
- 6: em\_estpl.\*
- 7: em\_estro.\*
- 8: em\_estri.\*
- 9: em\_estti.\*
- 10: gb\_est2.\*
- 11: gb\_est3.\*
- 12: gb\_gss.\*
- 13: em\_gss\_hum.\*
- 14: em\_gss\_inv.\*
- 15: em\_gss\_pln.\*
- 16: em\_gss\_vrt.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Query	Score	Match	Length	ID	Description
c 1	70	45.8	593	9	AL659353	AL659353
2	70	45.8	645	9	AL644594	AL644594
3	67.4	44.1	619	9	BB155356	BB155356
4	67.4	44.1	660	9	BB461065	BB461065
5	62.8	41.0	430	9	AL601246	AL601246
6	62.8	41.0	537	9	AL890775	AL890775
7	62.8	41.0	821	10	BF239967	BF239967
8	62.8	41.0	866	9	AU125712	AU125712
9	62.8	41.0	1028	9	BE895133	BE895133
10	61	39.9	446	9	BB834922	BB834922
c 11	55.8	36.5	686	9	AW997058	AW997058
12	54.8	35.8	438	9	BB830730	BB830730
c 13	54.8	35.8	856	12	CNS04DVG	AL286261
c 14	47.8	31.2	532	10	BI844947	BI844947
c 15	47.8	31.2	574	10	BM185888	BM185888
c 16	47.8	31.2	578	10	BM070535	BM070535
c 17	47.8	31.2	608	10	BM024825	BM024825

c 18	47.8	31.2	608	10	BM185005	BM185005
c 19	47	30.7	337	9	AW996787	AW996787
c 20	45.2	29.5	1013	12	CNS055XK	AL322625
c 21	44.6	29.2	1156	12	CNS02JUV	AL200704
c 22	43	28.1	724	12	CNS05DSJ	AL32812
c 23	43	28.1	734	12	CNS056MU	AL323535
c 24	43	28.1	923	12	CNS04J5I	AL293103
c 25	43	28.1	924	12	CNS04I26	AL292875
c 26	43	28.1	963	12	CNS01X68	AL171305
c 27	43	28.1	1015	12	CNS05KJP	AL341950
c 28	43	28.1	1047	12	CNS04JPV	AL293839
c 29	43	28.1	1050	12	CNS05N3J	AL344872
c 30	43	28.1	1054	12	CNS05O6W	AL346289
c 31	43	28.1	1074	12	CNS05IOP	AL317122
c 32	43	28.1	1096	12	CNS05IC3	AL316668
c 33	41.8	27.3	1017	12	CNS033G4	AL226093
c 34	41.8	27.3	1122	12	CNS05T9J	AL352864
c 35	41.4	27.1	967	12	CNS02BFS	AL189793
c 36	41.4	27.1	1029	12	CNS05GSM	AL336703
c 37	41.4	27.1	1036	12	CNS04VNY	AL309319
c 38	41.4	27.1	1038	12	CNS05AVS	AL329041
c 39	41.4	27.1	1051	12	CNS02HG6	AL197249
c 40	41.4	27.1	1056	12	CNS05R9Z	AL350288
c 41	41.4	27.1	1079	12	CNS05HIG	AL337633
c 42	41.4	27.1	1080	12	CNS054WU	AL321303
c 43	41.4	27.1	1101	12	CNS04UQA	AL328107
c 44	41.4	27.1	1101	12	CNS04WXH	AL310958
c 45	41.4	27.1	1101	12	CNS0592M	AL327883

#### ALIGNMENTS

RESULT 1  
AL659353/c  
LOCUS  
DEFINITION  
AL659353 XGC-neurula Silurana tropicalis cDNA clone Tneu045e20 5', mRNA linear EST 13-DEC-2001  
mRNA sequence.  
ACCESSION  
AL659353  
VERSION  
AL659353.1 GI:17672995  
KEYWORDS  
EST.  
SOURCE  
western clawed frog.  
ORGANISM  
Silurana tropicalis  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;  
Xenopodinae; Silurana.  
REFERENCE  
1 (bases 1 to 593)  
Huckle,E., Taylor,R., Ashurst,J.L., Zorn,A.M. and Rogers,J.  
Sanger Xenopus tropicalis EST project 2001 (10\_2001)  
TITLE  
Unpublished (2001)  
JOURNAL  
Contact: Huckle E  
COMMENT  
Sanger Centre  
Hinnton, Cambridgeshire, CB10 1SA, UK  
Email: trop@sanger.ac.uk  
Sanger Xenopus tropicalis EST project 2001  
TROPICALIS SEQUENCE ID: Tneu045e20.sp6  
Sequencing primer: SP6  
This sequence is from a Xenopus Gene Collection (XGC) library  
Constructed by Aaron M. Zorn.  
Location/Qualifiers  
1..593  
/organism="Silurana tropicalis"  
/db\_xref="taxon:8364"  
/clone="Tneu045e20"  
/dev\_stage="neurula"  
/lab\_host="Escherichia coli DH10B"  
/note="Vector: pCS107; Site1: EcoRI; Site2: NotI; cDNA was oligo dT primed from Sug of poly A+ RNA from neurula. EcoRI-NotI cut cDNA was then ligated into pCS107 with EcoRI at the 5' end and NotI at the 3' end."

BASE COUNT  
119 a 134 c 114 g 226 t  
ORIGIN



sequencing pipeline with 384 multiplexed sequencer. Genome Res. 10 (10), 1757-1771 (2000)

Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y., and Hayashizaki, Y.

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

Kondo, S., Shinagawa, A., Saito, T., Kiyosawa, H., Yamanaka, I., Aizawa, K., Fukuda, S., Harada, A., Itoh, M., Kawai, Y., Shibata, K., and Hayashizaki, Y.

Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)

Please visit our web site (<http://genome.gsc.riken.go.jp/>) for further details.

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN, Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

```

case: radsdes:
  location/Qualifiers
  id: 5660
  /orgism="Mus musculus"
  /abstraf="axon:10090"
  /clone="D130070B13"
  /clone.lib="P18EN full-length enriched, 12 days embryo
    spinal ganglion"
  /tissue_type="spinal ganglion"
  /dev_stage="12 days embryo"
  /lab_host="DH10B"

```

/note=Site.1: Sall; Site.2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN, Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer 15'.

GAGAGAGAGCGGCGCACTCGAGTTTTTTTTTTTTTN 3'), cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adaptor of sequence [5',

GAGACAGAGATTCGACGACTTAATAATATCCCCCCCCCCC 3']. cDNA  
was cloned with BamHI and XhoI. Vector: a modified  
pBluescript KS(+) after bulk excision from Lambda FLC 1.  
a 130 c 151 t 158 g 151 t

44.13: Score 67.4; DB 9; Length 660;  
ty 79.2%; Pred. No. 1.8e-08;  
servative 0; Mismatches 21; Indels 0; Gaps 0;

AGGCAAGACGCTACAGACCCGTCAGACTACCTCATTTAAATTACTGAAT 111  
 AAGCAAAACAGCTTACAGACCCGTCAGACTACCTCATCAAACTACTTAGC 561  
 CAGAAAGGGAAGCACAAGCGCTTCGCTGGTGC 152  
 CAAAGAGAGCGCTACAGACCTTTGTGGTGC 602

430 bp linear mRNA EST 14-AUG-2001  
U11040.rl.313 (synonym: h1ccc2) Homo sapiens cDNA clone  
U11040.5', mRNA sequence.

1 GI:15164752

Mammalia: Eutheria; Primates; Catarrhini; Hominoidea: Hominae: Hominidae: Homo.

REFERENCE	1 (bases 1 to 430)
AUTHORS	Bloecker, H., Boecher, M., Brandt, P., Mewes, W., Weil, B. and Wiemann, S.
TITLE	EST (Bloecker, H., Boecher, M., Brandt, P., Mewes, H. W., Weil, B. and Wiemann, S.)
JOURNAL	Unpublished (1999)
COMMENT	Contact: Bloecker H MPS Am Klopferspitz 18a D-82152 Martinsried, Germany This is the 5' sequence of the clone insert Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de; sequenced by GBF (National Research Centre for Biotechnology Ltd., Braunschweig/Germany) within the cDNA sequencing consortium of the German Genome Project. No SI sequence available. This clone (DKFZp313J1040) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de

[illegible]

```

www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 1924 Std Error: 0.00
Seq primer: -400P from Gibco
High quality sequence stop: 418.
      Location/Qualifiers
1..547
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2443725"
/clone_lib="NCI_CGAP_Ut2"
/tissue_type="moderately-differentiated endometrial
adenocarcinoma, 3 pooled tumors"
/lab_host="DH10B"
/notes="Organ: uterus; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dr.
Average insert size 1.85 kb. Life Technologies catalog #:
11539-012"
114 a 118 c 85 g 230 t
BASE COUNT
ORIGIN

```

## FEATURES

```

1. 821
   /organism="Homo sapiens"
   /db_xref="taxon:9606"
   /clone="IMAGE:4133129"
   /clone_lib="NIH_MGC_54"
   /tissue_type="from Chronic myelogenous leukemia"
   /lab_host="DH10B (T1 phage-resistant)"
   /note="Organ: bone marrow; Vector: pDNR-LIB (Clontech);
   site_1: Sf11 (gcccctatggcc); Site_2: Sf11 (gcccctatggcc
   ); Double-stranded cDNA was prepared from cell line RNA.
   5' and 3' adaptors were used in cloning as follows: 5'

```

adaptor sequence: 5'-CACGGCCATTATGCC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCGGAGCGCCGATG-dt(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.75 kb (range 0.9-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA)."

BASE COUNT 345 a 149 c 190 g 137 t  
ORIGIN

Query Match 41.0%; Score 62.8; DB 10; Length 821;  
Best Local Similarity 77.6%; Pred. No. 3.5e-07;  
Matches 76; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

QY 52 AAGAAACCCAGGCAAGACGCTACAGACCGGTGCAGACTACCTCATTAATTAAGTGAAT 111  
DB 9 AAAAAACCAAGCAAAACAGTTGACAGACCGGTGCAGACTACCTCATCAAAATTAAGTGAAT 68

QY 112 AAAGACCTTGCAGAAAGGAGGACACAAAGGCTTGCTGG 149  
DB 69 AGATCTTGCAGAAAGGAGGCTTCTTCTGCTGGCGG 106

RESULT 8  
LOCUS AUI25712 866 bp mRNA linear EST 23-OCT-2000  
DEFINITION AUI25712 NT2RM4 Homo sapiens cDNA clone NT2RM4002061 5', mRNA  
sequence.  
ACCESSION AUI25712  
VERSION AUI25712.1 GI:10950428  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 866)  
Ota.T., Wakamatsu.A., Ozawa.M., Ishii.S., Saito.K., Yamamoto.J., Nakamura.Y., Nishikawa.T., Nagai.T., Suzuki.Y., Sugano.S. and Isogai.T.  
HRI human cDNA project (Ota.T., Wakamatsu.A., Ozawa.M., Ishii.S., Saito.K., Yamamoto.J., Nakamura.Y., Nishikawa.T., Nagai.T., Suzuki.Y., Sugano.S., Isogai.T.)  
Unpublished (2000)  
CONTACT: Takao Isogai  
Genomics Laboratory  
Helix Research Institute  
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan  
Tel: 81-438-52-3951  
Fax: 81-438-52-3952  
Email: genomics@hri.co.jp  
HRI human cDNA project: 5'- & 3'-end one pass sequencing: Helix Research Institute; cDNA library construction: Department of Virology, Institute of Medical Science, University of Tokyo, and Helix Research Institute.

FEATURES  
source  
1..866  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="NT2RM4002061"  
/clone\_lib="NT2RM4"  
/cell\_type="teratocarcinoma"  
/cell\_line="NT2"  
/notes="Vector: pME18SFL3; mRNA from uninduced NT2 neuronal precursor cells"

BASE COUNT 312 a 149 c 196 g 207 t  
ORIGIN

Query Match 41.0%; Score 62.8; DB 9; Length 866;  
Best Local Similarity 77.6%; Pred. No. 3.5e-07;  
Matches 76; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

QY 52 AAGAAACCCAGGCAAGACGCTACAGACCGGTGCAGACTACCTCATTAATTAAGTGAAT 111  
DB 471 AAAAAACCAAGCAAAACAGTTGACAGACCGGTGCAGACTACCTCATCAAAATTAAGTGAAT 530

QY 112 AAAGACCTTGCAGAAAGGAGGACACAAAGGCTTGCTGG 149  
DB 531 AGATCTTGCAGAAAGGAGGCTTCTTCTGCTGGCGG 568

RESULT 9  
LOCUS BB895133 1028 bp mRNA linear EST 20-OCT-2000  
DEFINITION BB895133 601436060F1 NIH\_MGC\_72 Homo sapiens cDNA clone IMAGE:3921087 5', mRNA sequence.  
ACCESSION BB895133  
VERSION BB895133.1 GI:10358221  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 1028)  
NIH-MGC http://mgs.nci.nih.gov/.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
CONTACT: Robert Strausberg, Ph.D.  
Email: cgapbs@mail.nih.gov  
Tissue Procurement: ATCC/DCTD/OPP  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLAW9753 row: h column: 16  
High quality sequence stop: 488.

FEATURES  
source  
1..1028  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3921087"  
/clone\_lib="NIH\_MGC\_72"  
/tissue\_type="melanotic melanoma"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: skin; Vector: pCMV-SPORT6; Site: 1: NotI; Site: 2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 2 kb. Library constructed by Life Technologies."

BASE COUNT 387 a 205 c 238 g 198 t  
ORIGIN

Query Match 41.0%; Score 62.8; DB 10; Length 1028;  
Best Local Similarity 77.8%; Pred. No. 3.6e-07;  
Matches 76; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

QY 52 AAGAAACCCAGGCAAGACGCTACAGACCGGTGCAGACTACCTCATTAATTAAGTGAAT 111  
DB 152 AAAAAACCAAGCAAAACAGTTGACAGACCGGTGCAGACTACCTCATCAAAATTAAGTGAAT 211

QY 112 AAAGACCTTGCAGAAAGGAGGACACAAAGGCTTGCTGG 149  
DB 212 AGATCTTGCAGAAAGGAGGCTTCTTCTGCTGGCGG 249

RESULT 10  
LOCUS BB834922 446 bp mRNA linear EST 19-NOV-2001  
DEFINITION BB834922 RIKEN full-length enriched, mammary gland RCB-0527 JYG-MC(B) cDNA Mus musculus cDNA clone G93003J21 3', mRNA sequence.  
ACCESSION BB834922  
VERSION BB834922.1 GI:17013165







Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes  
 i Cyprinidae; Danio.  
 1 (bases 1 to 574)  
 Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy  
 S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood  
 K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B.,  
 Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E.,  
 Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.  
 and Wilson,R.  
 WashU Zebrafish EST Project 1998  
 Unpublished (1998)  
 Other ESTs: fv74a09.y1  
 Contact: Stephen L. Johnson  
 Washington University School  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: zbrfish@watson.wustl.edu  
 cDNA Library Preparation: John Ngai. cDNA Library Arrayed by:  
 Matthew Clark. DNA Sequencing by: Washington University Genome  
 Sequencing Center Clone distribution: Genome Systems, St. Louis,  
 Missouri (web address: [www.genomesystems.com](http://www.genomesystems.com)) (email contact:  
[info@genomesystems.com](mailto:info@genomesystems.com)) and Research Genetics, Huntsville, Alabama  
 (web address: [www.resgen.com](http://www.resgen.com)) (email contact: [info@resgen.com](mailto:info@resgen.com)) and  
 RessourcenzentrumPrimatendatenbank, Berlin, Germany (web address:  
[www.rzpd.de](http://www.rzpd.de))

```

Seq primer: T7 from Gibco
High quality sequence stop: 328.
Location/Qualifiers
1..574
/organism="Danio rerio"
/db_xref="taxon:7955"
/clone="5413937"
/clone.lib="zebrafish adult brain"
/sex="mixed male and female"
/tissue_type="brain"
/dev_stage="adult"
/lab_host="E. coli DH10B"
/note="vector: pZiPLOX; Site_1: NotR; Site_2: SalI;
Original library was constructed in lambdaZIPLOX. Mass
excision of the cDNA library was performed to yield
pZiPLOX plasmids. Insert check was done in original
library."
100 a 116 a 223 +
135 c

```

	Query Match	31.2%;	Score 47.8;	DB 10;	Length 574;	
	Best Local Similarity	73.5%;	Pred. No. 0.0042;			
Matches	61;	Conservative	0;	Mismatches 22;	Indels 0;	Gaps 0;
52	AGAAACCCAGGCAAGCAGCTACAGACCCGTCAGACTACCTCATTAATAACTGAAT					
349	AGAAGACCTCAAGGAAGCAGTTACAGATCCGAACAGACTACTTGCTGAAGATGCTCAAG					
112	AAGACCTTGCAGAAAGGAAGC					
289	AAGACCTGGAAGCAAGATGC					

Search completed: August 2, 2002, 16:52:14  
Job time: 12215 sec

Mon Aug 5 11:51:51 2002

us-08-973-363-3.rst

Page 9

us-08-973-363-3.rst

**THIS PAGE BLANK (USPTO)**

OM nucleic - nucleic search, using sw model

Run on: August 2, 2002, 20:47:20 ; Search time 1205.34 seconds  
(without alignments)  
217.937 Million cell updates/sec

Title: US-08-973-363-4  
 Perfect score: 153  
 Sequence: 1 ATTTTACCTGATGATCCAGA.....CACAGAGACTGCTGGTGCA 153

Scoring table: IDENTITY\_NUC  
 Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0  
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
 Maximum Match 100%  
 Listing first 45 summaries

Database : N.Geneseq\_032802.\*

1:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1980.DAT.*
2:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1981.DAT.*
3:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1982.DAT.*
4:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1983.DAT.*
5:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1984.DAT.*
6:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1985.DAT.*
7:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1986.DAT.*
8:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1987.DAT.*
9:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1988.DAT.*
10:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1989.DAT.*
11:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1990.DAT.*
12:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1991.DAT.*
13:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1992.DAT.*
14:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1993.DAT.*
15:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1994.DAT.*
16:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1995.DAT.*
17:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1996.DAT.*
18:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1997.DAT.*
19:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT.*
20:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT.*
21:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT.*
22:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT.*
23:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT.*
24:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	153	100.0	153	AA142758	Chick CHD-W gene f
2	140.2	91.6	153	AA142757	Chick CHD-1A gene
3	135.4	86.5	153	AA142759	Great tit CHD-W ge
4	103.6	69.0	153	AA142756	Mouse CHD-1 gene (
5	100.4	65.6	1316	AA142754	Chicken CHD-W gene
6	95	62.1	6608	AA142751	Chicken CHD-1A gen
7	41.2	26.9	1311	AA142752	Altered telomere r
8	39.6	25.9	421	AAK88882	Human digestive sy
9	39.6	25.9	421	AA157603	Human colorectal c

## ALIGNMENTS

RESULT 1

ID	AA142758	standard; DNA; 153 BP.
AC	AA142758;	
DT	12-MAR-1997 (first entry)	
DE	Chick CHD-W gene fragment.	
XX	Bird; sex determination; chromodomain-Helicase-DNA binding 1;	
KW	CHD-1A; CHD-W; W chromosome; ss.	
XX	Gallus sp.	
XX	Key	Location/Qualifiers
FT	misc_difference 52..81	
FT	/tag= a	
FT	/note= "bases 52-81 are a repeat of bases 22-51 and are ignored in the translated amino acid sequence given in Fig 3"	
XX	WO9639505-A1.	
PD	12-DEC-1996.	
XX	05-JUN-1996; 96WO-GB01341.	
XX	06-JUN-1995; 95GB-0011439.	
XX	(ISIS-) ISIS INNOVATION LTD.	



XX Avian chromodomain-helicase-DNA binding genes determine sex in  
 PT birds - used for sex determ. and to control sex of progeny  
 XX  
 PS Claim 8; Fig 3; 76pp; English.  
 XX Bases 3855-3977 (AAT42756) of the mouse CHD-1 gene show homology  
 CC to portions of the chicken CHD-1A (A = Avian) gene (see also  
 CC and AAT42757), chicken CHD-W (W refers to the W chromosome) gene  
 CC and the great tit CHD-W gene (see also AAT42759).  
 CC Translated amino acid sequences of this region are provided in  
 CC sequences of this region are provided in AAW08146-49. The CHD-1A  
 CC (see also AAT42751) and CHD-W (see also AAT42754-55) genes determine  
 CC sex in birds and can be used to identify the sex of an embryo,  
 CC foetus etc. and to manipulate the sex of progeny.  
 XX  
 SQ Sequence 153 BP; 60 A; 33 C; 31 G; 29 T; 0 other;

Query Match 88.5%; Score 135.4; DB 18; Length 153;  
 Best Local Similarity 92.8%; Pred. No. 2.7e-34;  
 Matches 142; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 1 ATTTTACCTGATGCCAGATAGAAACCCAGGCTAAGCATTACAGACCAAGAAACCC 60  
 Db 1 attttacctgatgccagataagaacccagggcaagcaggttgcagaccagaaccca 60

QY 61 CAGGCTAAGCAGTTACAGACCCGTCGAGATTACCTCATTTAAATTAATTAAGACCTT 120  
 Db 61 caggcaaacaggttgcagaccgctgcagattacctcattataattactgaataaagacctt 120

QY 121 GCAGAGAGGAGGACAGACACTTCTGCTGCA 153  
 Db 121 gcaagaaagaagtgcagaaagacttactctgtgca 153

## RESULT 4

AAT42756  
 ID AAT42756 standard; DNA; 153 BP.  
 AC AAT42756;  
 DT 12-MAR-1997 (first entry)  
 DE Mouse CHD-1 gene (bases 3855-977).  
 XX Bird; sex determination; chromodomain-Helicase-DNA binding 1;  
 KW CHD-1; CHD-W; W chromosome; ss.  
 XX Mus sp.  
 OS  
 XX Key Location/Qualifiers  
 FH misc\_difference 52..81  
 ET /\*tag- a  
 FT /note- "bases 52-81 are a repeat of bases 22-51  
 ET and are ignored in the translated amino  
 FT acid sequence given in Fig 3".  
 XX  
 PW WO9639505-A1.  
 XX  
 PD 12-DEC-1996.  
 XX  
 PF 05-JUN-1996; 96WO-GB01341.  
 XX  
 PR 06-JUN-1995; 95GB-0011439.  
 XX  
 PA (ISIS-) ISIS INNOVATION LTD.  
 XX  
 PI Griffiths R, Tiwari B;  
 XX WPI; 1997-043127/04.  
 DR P-PSDB; AAW08146.  
 XX Avian chromodomain-helicase-DNA binding genes determine sex in

PT birds - used for sex determ. and to control sex of progeny  
 XX  
 PS Claim 8; Fig 3; 76pp; English.  
 XX Bases 3855-3977 (AAT42756) of the mouse CHD-1 gene show homology  
 CC to portions of the chicken CHD-1A (A = Avian) gene (see also  
 CC and AAT42757), chicken CHD-W (W refers to the W chromosome) gene  
 CC (see also AAT42758) and the great tit CHD-W gene (see also AAT42759).  
 CC Translated amino acid sequences of this region are provided in  
 CC sequences of this region are provided in AAW08146-49. The CHD-1A  
 CC (see also AAT42751) and CHD-W (see also AAT42754-55) genes determine  
 CC sex in birds and can be used to identify the sex of an embryo,  
 CC foetus etc. and to manipulate the sex of progeny.  
 XX  
 SQ Sequence 153 BP; 61 A; 37 C; 27 G; 28 T; 0 other;

Query Match 69.0%; Score 105.6; DB 18; Length 153;  
 Best Local Similarity 80.9%; Pred. No. 1.1e-24;  
 Matches 123; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 1 ATTTTACCTGATGCCAGATAGAAACCCAGGCTAAGCATTACAGACCAAGAAACCC 60  
 Db 1 attttacctgatgccagataagaacccagggcaagcaggttgcagaccagaaccca 60

QY 61 CAGGCTAAGCAGTTACAGACCCGTCGAGATTACCTCATTTAAATTAATTAAGACCTT 120  
 Db 61 caagcaaacaggttgcagaccgctgcagattacctcattataattactgaataaagacctt 120

QY 121 GCAGAGAGGAGGACAGACACTTCTGCTGTC 152  
 Db 121 gcaaaaaagagaggctcagagacttctgctgctc 152

## RESULT 5

AAT42754  
 ID AAT42754 standard; cDNA; 1316 BP.  
 AC AAT42754;  
 DT 12-MAR-1997 (first entry)  
 DE Chicken CHD-W gene (partial sequence).  
 XX Bird; sex determination; chromodomain-Helicase-DNA binding 1 Avian;  
 KW CHD-W; chromodomain-Helicase-DNA binding on the W chromosome; ss.  
 XX Gallus sp.  
 XX WO9639505-A1.  
 XX 12-DEC-1996.  
 XX 05-JUN-1996; 96WO-GB01341.  
 XX 06-JUN-1995; 95GB-0011439.  
 XX (ISIS-) ISIS INNOVATION LTD.  
 XX Griffiths R, Tiwari B;  
 XX WPI; 1997-043127/04.  
 XX Avian chromodomain-helicase-DNA binding genes determine sex in  
 PT birds - used for sex determ. and to control sex of progeny  
 XX  
 PS Claim 1; Fig 8; 76pp; English.

The chicken CHD-W gene (AAT42754) acting alone or in conjunction with  
 CC the closely related CHD-1A gene (AAT42751) is suggested to initiate  
 CC female development in birds. The sequence of CHD-W was deduced  
 CC from 2 clones isolated from a 10-day chick embryo library using  
 CC a fragment of the CHD-1A gene as a probe. The CHD-W gene is

CC located on the W chromosome. Probes based on CHD-W and CHD-1A give  
CC a W chromosome-specific signal on hybridisation to genomic DNA of a  
CC non-rare bird and can be used for sex determination of a bird. CHD-W  
CC nucleic acids can also be used to control the sex of progeny of a  
CC bird.

XX SQ Sequence 1316 BP; 492 A; 208 C; 306 G; 304 T; 6 other;

Query Match 65.68; Score 100.4; DB 18; Length 1316;  
Best Local Similarity 99.08; Pred. No. 1.1e-22;  
Matches 101; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 52 AAGAAACCCAGGCTAAGCAGTTACAGACCGTGCAGATTACCTCATTAATAATTAAGTGAAT 111  
|||||

Db 1159 aagaacccccaggctaaagcagttacagactcgtgcagattacaccttaataattactgaat 1218  
|||||

Qy 112 AAGACCTTGCAAGAAAGACGACAGACTTGTGTGCA 153  
|||||

Db 1219 aaagaccttgaagaagaagacagacagacttgcgtgtgca 1260  
|||||

## RESULT 6

AAT42751  
ID AAT42751 standard; cDNA; 6608 BP.

XX AC AAT42751;

XX DT 12-MAR-1997 (first entry)

XX DE Chicken CHD-1A gene.

XX KW Bird; sex determination; Chromodomain-Helicase-DNA binding 1 Avian;

XX KW CHD-W; chromodomain-Helicase-DNA binding on the W chromosome; ss.

XX OS Gallus sp.

XX FH Key Location/Qualifiers

XX FT CDS 228..5390

XX FT /\*tag= a

XX PN WO9639505-A1.

XX XX 12-DEC-1996.

XX XX 05-JUN-1996; 96WO-GB01341.

XX XX 06-JUN-1995; 95GB-0011439.

XX XX (ISIS-) ISIS INNOVATION LTD.

XX XX Griffiths R, Tiwari B;

XX XX WPI; 1997-043127/04.

XX XX Avian chromodomain-helicase-DNA binding genes determine sex in

XX XX birds - used for sex determin. and to control sex of progeny

XX XX Claim 1; Fig 5; 76pp; English.

XX CC The chicken CHD-W gene (AAT42754) acting alone or in conjunction with  
XX CC the closely related CHD-1A gene (AAT42751) is suggested to initiate  
XX CC female development in birds. The sequence of CHD-1A was deduced  
XX CC from 3 clones isolated from a stage 10-12 chicken embryo cDNA  
XX CC library using a great tit CHD-W sequence (see also AAT42755) as probe.  
XX CC The CHD-1A (A = Avian) gene shows close identity to the mouse CHD-1  
XX CC gene (see also AAT42756-57). It is located on an autosome or Z  
XX CC chromosome. Probes based on CHD-W and CHD-1A give a W chromosome-  
XX CC specific signal on hybridisation to genomic DNA of a non-rare  
XX CC bird and can be used for sex determination of a bird. CHD-1A nucleic  
XX CC acids can also be used to control the sex of the progeny of a bird.

XX SQ Sequence 6608 BP; 2289 A; 1207 C; 1459 G; 1653 T; 0 other;

Query Match 62.18; Score 95; DB 18; Length 6608;  
Best Local Similarity 95.18; Pred. No. 1.1e-20;  
Matches 98; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 51 CAAGAAACCCAGGCTAAGCAGTTACAGACCGTGCAGATTACCTCATTAATAATTAAGTGA 110  
|||||

Db 4100 caagaaacccccaggcgaagcagctacagaccgctgcagactacattactgaat 4159  
|||||

Qy 111 TAAGACCTTGCAAGAAAGACGACAGACTTGTGTGCA 153  
|||||

Db 4160 taagaccttgcagaagaaggaagcacaaggtgtgtgtgca 4202  
|||||

## RESULT 7

AAV59280  
ID AAV59280 standard; cDNA; 1311 BP.

XX AC AAV59280;

XX DT 14-DEC-1998 (first entry)

XX DE Altered telomere repeat binding factor 1 gene.

XX KW ds; human; telomere repeat binding factor; A-TRF; dimerisation domain;

XX KW telomere; ageing; ataxia telangeiectasia; Down's syndrome; tumour; viral.

XX OS Homo sapiens.

XX OS Synthetic.

XX FH Key Location/Qualifiers

XX FT CDS 1..1311

XX FT /\*tag= a

XX FT /product= "A-TRF"

XX XX WO9836066-A1.

XX XX 20-AUG-1998.

XX XX 13-FEB-1998; 98WO-US02765.

XX XX 04-FEB-1998; 98US-0018628.

XX XX 13-FEB-1997; 97US-0800264.

XX XX (UYRQ ) UNIV ROCKEFELLER.

XX XX Bianchi A, De Lange T, Van Steensel B;

XX XX WPI; 1998-480769/41.

XX XX P-PSDB; AAV59280.

XX PT Nucleic acid encoding altered telomere repeat binding protein and  
XX PT related vectors - transformants, hetero-dimers and antibodies, used  
XX PT to inhibit shortening of telomerases caused by ageing or disease,  
XX PT also used to extend life of cells in culture

XX PS Claim 14; Page 110-111; 163pp; English.

XX CC The altered vertebrate telomere repeat binding protein (A-TRF) has a  
XX CC telomere repeat binding factor (TRF) dimerisation domain, and forms a  
XX CC hetero-dimer with TRF, preventing it from binding to the specified repeat  
XX CC sequence. A-TRF, optionally expressed by gene therapy, is used to  
XX CC inhibit shortening of telomerases associated with ageing (for cosmetic  
XX CC purposes) and disease, e.g. ataxia telangeiectasia, Down's syndrome,  
XX CC atrophy of the skin, age-related macular degeneration, atherosclerosis,  
XX CC tumours and viral (including human immune deficiency virus) infection.  
XX CC Cells expressing A-TRF also have an increased life span in vitro, e.g.  
XX CC for expression of recombinant proteins or where intended for subsequent  
XX CC transplant or for testing, eliminating the need for transformation.

XX SQ Sequence 1311 BP; 441 A; 257 C; 316 G; 297 T; 0 other;



	Query Match	26.9%	Score 41.2;	DB 19;	Length 1311;
	Best Local Similarity	85.2%;	Pred. No. 0.0015;		
	Matches	46;	Conservative	0;	Mismatches 8; Indels 0; Gaps 0;
OY	52	AAGAAACCCCGGTTAGCAGTTACAGACCCGTCGAGATTACCTCAATTA	105		
Db	1258	aaagaaccacgaacaaacagttatcaacacccgtgcagactacctcatcaacta	1311		

RESULT	8
AAK8882	
ID	AAK8882 standard; CDNA; 421 BP.
XX	
XX	AAK8882;
AC	
XX	
DT	05-NOV-2001 (first entry)
DT	
DE	Human digestive system antigen coding sequence SEQ ID NO: 1198.
DE	
XX	Human; digestive system antigen; gene therapy; cancer; appendicitis;
KW	ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;
KW	digestive system disorder; Meckel's diverticulum; ss.

AA Homo sapiens.

WO200155314-A2.

02-AUG-2001.

17-JAN-2001:

31-JAN-2000: 2000US-0179065.

04 - FEB - 2000; 2000US-0180628;  
PK  
24 - FEB - 2000; 2000US-0184664;  
PB

PR 02-MAR-2000; 2000US-0186330.  
PR 16-MAR-2000; 2000US-0189874

PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.

PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467

PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135

PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880

PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496

PR 14-JUL-2000; 2000DS-0218290.  
PB 26-JUL-2000; 2000DS-0230053

26-JUL-2000; 2000US-0220964.  
14-AUG-2000; 2000US-0324518

PR 14-AUG-2000; 2000US-0224519.  
DB 14-AUG-2000; 2000US-0224519.

PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225214.

PR 14-AUG-2000; 2000US-0225267.

PR 14-AUG-2000; 2000US-0225270.

PR 14-AUG-2000; 2000US-0225757.

PR 14-AUG-2000; 2000US-0225759.

PR 22-AUG-2000; 2000US-0226681.

PR 22-AUG-2000; 2000US-0227182.

PR 30-AUG-2000; 2000US-0228924.

PR 01-SEP-2000; 2000US-0229343.

PR 01-SEP-2000; 2000US-0229345.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
84

PR	05-SEP-2000;	2000US-02235513;
PR	06-SEP-2000;	2000US-02304337;
PR	06-SEP-2000;	2000US-02304336;
PR	08-SEP-2000;	2000US-02311242;
PR	08-SEP-2000;	2000US-02311243;
PR	08-SEP-2000;	2000US-02311244;
PR	08-SEP-2000;	2000US-02311413;
PR	08-SEP-2000;	2000US-02311414;
PR	08-SEP-2000;	2000US-02311418;
PR	08-SEP-2000;	2000US-02320811;
PR	12-SEP-2000;	2000US-02319681;
PR	14-SEP-2000;	2000US-02323397;
PR	14-SEP-2000;	2000US-02323398;
PR	14-SEP-2000;	2000US-02323399;
PR	14-SEP-2000;	2000US-02324040;
PR	14-SEP-2000;	2000US-02324041;
PR	14-SEP-2000;	2000US-02330633;
PR	14-SEP-2000;	2000US-02330634;
PR	21-SEP-2000;	2000US-02344223;
PR	21-SEP-2000;	2000US-02344277;
PR	25-SEP-2000;	2000US-02349597;
PR	25-SEP-2000;	2000US-02349598;
PR	26-SEP-2000;	2000US-02354844;
PR	27-SEP-2000;	2000US-02355834;
PR	27-SEP-2000;	2000US-02358536;
PR	29-SEP-2000;	2000US-02363217;
PR	29-SEP-2000;	2000US-02363667;
PR	29-SEP-2000;	2000US-02363668;
PR	29-SEP-2000;	2000US-02363669;
PR	02-OCT-2000;	2000US-02365070;
PR	02-OCT-2000;	2000US-02368602;
PR	02-OCT-2000;	2000US-02370337;
PR	02-OCT-2000;	2000US-02370738;
PR	02-OCT-2000;	2000US-02370739;
PR	02-OCT-2000;	2000US-02370740;
PR	02-OCT-2000;	2000US-02417885;
PR	02-OCT-2000;	2000US-02417887;
PR	13-OCT-2000;	2000US-02399933;
PR	13-OCT-2000;	2000US-02399937;
PR	20-OCT-2000;	2000US-02409660;
PR	20-OCT-2000;	2000US-02412121;
PR	20-OCT-2000;	2000US-02417865;
PR	20-OCT-2000;	2000US-02417866;
PR	20-OCT-2000;	2000US-02418008;
PR	20-OCT-2000;	2000US-02418009;
PR	20-OCT-2000;	2000US-02418266;
PR	01-NOV-2000;	2000US-02461474;
PR	08-NOV-2000;	2000US-02464675;
PR	08-NOV-2000;	2000US-02464676;
PR	08-NOV-2000;	2000US-02464677;
PR	08-NOV-2000;	2000US-02464678;
PR	08-NOV-2000;	2000US-02465233;
PR	08-NOV-2000;	2000US-02465234;
PR	08-NOV-2000;	2000US-02465235;
PR	08-NOV-2000;	2000US-02465237;
PR	08-NOV-2000;	2000US-02465238;
PR	08-NOV-2000;	2000US-02465239;
PR	08-NOV-2000;	2000US-02466010;
PR	08-NOV-2000;	2000US-02466019;
PR	08-NOV-2000;	2000US-02466020;
PR	08-NOV-2000;	2000US-02466133;
PR	17-NOV-2000;	2000US-02466134;
PR	17-NOV-2000;	2000US-02492007;
PR	17-NOV-2000;	2000US-02492008;
PR	17-NOV-2000;	2000US-02492010;
PR	17-NOV-2000;	2000US-02492019;
PR	17-NOV-2000;	2000US-02492112;
PR	17-NOV-2000;	2000US-02492113;
PR	17-NOV-2000;	2000US-02492114;
PR	17-NOV-2000;	2000US-02492115;
PR	17-NOV-2000;	2000US-02492116;
PR	17-NOV-2000;	2000US-02492117;

```
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 06-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-502630/55.
DR P-PSDB; AAM93109.
XX
XX Polynucleotides encoding digestive system antigens, useful for
PT diagnosing, treating, preventing and/or prognosing disorders of the
PT digestive system, particularly cancer and cancer metastases -
XX
XX Claim 1; SEQ ID NO 1198; 986pp; English.
PS
XX
XX The present invention provides the protein and coding sequences of a
CC number of human digestive system antigens. These can be used in the
CC diagnosis, treatment and prevention of digestive system disorders,
CC including cancer, Meckel's diverticulum, bacterial or parasitic
CC infections, appendicitis, Hirschsprung's disease, chronic colitis or
CC ulcerative colitis. The present sequence is a cDNA encoding a digestive
CC system antigen of the invention.
XX
XX Sequence 421 BP; 148 A; 78 C; 127 G; 61 T; 7 other;
SQ

Query Match 25.9%; Score 39.6; DB 22; Length 421;
Best Local Similarity 63.8%; Pred. No. 0.0034;
Matches 60; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 52 AAGAAACCCAGCGTAAAGCAGTTACAGACCCGTGCAGATTACCTCATTAATTAAGTGAAT 111
Dd 131 aaaaagcttcagggaagcagcagcagccgagcgagattactgttgaagctgctcaga 190

QY 112 AAAGACCTTGAAGAAAGGAGCAGAGACTTG 145
Dd 191 aaggtctggagaagaaggggctgtgacaggtg 224

RESULT 9
AAI57603
ID AAI57603 standard; cDNA; 421 BP.
XX
XX AAI57603;
AC
XX
XX 19-OCT-2001 (first entry)
DT
XX
XX Human colorectal cancer antigen cDNA SEQ ID NO: 67.
DE
XX
XX Human; colorectal cancer; colorectal cancer antigen; gene therapy; ss.
KW
XX
XX Homo sapiens.
OS
```

```
XX
PN WO200155350-A1.
XX
XX 02-AUG-2001.
PD
XX
PF 17-JAN-2001; 2001WO-US01350.
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 08-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
```



XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins  
CC (ABB57737-ABB72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX SQ Sequence 6240 BP; 1675 A; 1682 C; 1760 G; 1123 T; 0 other;

Query Match 22.9%; Score 35; DB 23; Length 6240;  
Best Local Similarity 70.1%; Pred. No. 0.25;  
Matches 47; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY 55 AAACCCAGGCTAAGCAGATTACAGACCCGTCGAGATTACCTCATTAATAATCTAGTAATAA 114  
II IIIIIII IIIII IIIII IIIII IIIII III III III  
Db 4205 aagccccaggccaagcagctgcagcgcgtgccgagtagctctcgaagatcatcaagaag 4264

QY 115 GACCTTG 121  
II III  
Db 4265 aacgtgg 4271

RESULT 11  
ABL06442  
ID ABL06442 standard; cDNA; 9933 BP.  
XX ABL06442;  
XX 26-MAR-2002 (first entry)  
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 13808.  
XX Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical; gene; ss.  
XX Drosophila melanogaster.  
XX WO200171042-A2.  
XX 27-SEP-2001.  
XX 23-MAR-2001; 2001WO-US09231.  
XX 23-MAR-2000; 2000US-191637P.  
XX 11-JUL-2000; 2000US-0614150.  
XX (PEKE ) PE CORP NY.  
XX Venter JC, Adams M, Li PWD, Myers EW;  
XX WPI; 2001-656860/75.  
XX P-PSDB; ABB62339.  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX Claim 1; SEQ ID NO 13808; 21pp + Sequence Listing; English.  
XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins

CC (ABB57737-ABB72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX SQ Sequence 9933 BP; 2711 A; 2468 C; 2538 G; 2216 T; 0 other;

Query Match 22.9%; Score 35; DB 23; Length 9933;  
Best Local Similarity 70.1%; Pred. No. 0.3;  
Matches 47; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY 55 AAACCCAGGCTAAGCAGATTACAGACCCGTCGAGATTACCTCATTAATAATCTAGTAATAA 114  
II IIIIIII IIIII IIIII IIIII IIIII III III III  
Db 6745 aagccccaggccaagcagctgcagcgcgtgccgagtagctctcgaagatcatcaagaag 6804

QY 115 GACCTTG 121  
II III  
Db 6805 aacgtgg 6811

RESULT 12  
AAQ63954  
ID AAQ63954 standard; cDNA; 1239 BP.  
XX AAQ63954;  
XX 14-OCT-1994 (first entry)  
XX P-selectin ligand.  
XX P-selectin ligand; glycoprotein; inflammation;  
KW intercellular adhesion; placenta;  
KW alpha-1,3/alpha-1,4-fucosyltransferase; PACE SOL;  
KW paired basic amino acid converting enzyme; ss.  
XX Homo sapiens.  
XX Key Location/Qualifiers  
FH CDS 1..1239  
FT /\*\*tag= a  
XX WO9410309-A.  
XX 11-MAY-1994.  
XX 22-OCT-1993; 93WO-US10168.  
XX 23-OCT-1992; 92US-0965662.  
XX 26-AUG-1993; 93US-0112608.  
XX (GEMY ) GENETICS INST INC.  
XX Chang X, Larsen GR, Sako DS, Veldman GM;  
XX WPI; 1994-167466/20.  
XX P-PSDB; R-53966.  
XX DNA encoding P-selectin ligand glyco:protein - for use in  
PT treating inflammatory disease characterised by P-selectin  
PT mediated inter:cellular adhesion  
XX Disclosure; Page 50-52; 64pp; English.  
XX CDNA of sequence AAQ63953, encoding P-selectin ligand, was derived  
CC from promyelocyte line HL60 clone PMF21:PL85. The ligand comprises  
CC residues 1-402, 1-310 (sol.), 42-402 (mature) or 42-310 (sol.),  
CC mature) of sequence AAR53965. An allelic variant (AAQ63954, AAR53966)  
CC was derived from human placenta. Co-transfection of a mammalian host  
CC cell with DNA encoding residues 42-310, plus DNA encoding  
CC alpha-1,3/alpha-1,4-fucosyltransferase and DNA encoding a soluble  
CC form of a paired basic amino acid converting enzyme (PACE SOL),  
CC encoded by sequence AAQ63957, allowed production of sol. mature





Mon Aug 5 11:51:52 2002

us-08-973-363-4.rng

Page 11

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 2, 2002, 16:52:14 ; Search time 8940.62 Seconds  
(without alignments)  
230.972 Million cell updates/sec

Title: US-08-973-363-4  
Perfect score: 153  
Sequence: 1 ATTTTACCTGATGATCCAGA.....CACAGAGACTTCTGGTCCA 153

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_htc:\*  
12: gb\_ysc:\*  
13: gb\_ysc\_hum:\*  
14: em\_ysc\_inv:\*  
15: em\_ysc\_pin:\*  
16: em\_ysc\_vit:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	69	45.1	619	9	BB155356
2	69	45.1	660	9	BB461065
3	66.8	43.7	593	9	AL659353
4	66.8	43.7	645	9	AL644594
5	64.2	42.0	446	9	BB834922
6	61.2	40.0	430	9	AL601246
7	61.2	40.0	547	9	AI850775
8	61.2	40.0	821	10	BP239967
9	61.2	40.0	866	9	AU125712
10	61.2	40.0	1028	10	BB895133
11	56.4	36.9	438	9	BB830730
12	54.2	35.4	686	9	AW997058
13	53.2	34.8	856	12	CNS04DVG
14	46.2	30.2	532	10	BI844947
15	46.2	30.2	574	10	BM185888
16	46.2	30.2	578	10	BM070535
17	46.2	30.2	608	10	BM024825

C 18	46.2	30.2	608	10	BM185005
C 19	45.4	29.7	337	9	AW996787
C 20	45	29.4	642	9	AL658477
C 21	43.6	28.5	1013	12	CNS055XK
C 22	43	28.1	924	12	CNS04I26
C 23	43	28.1	1156	12	CNS02JUV
C 24	41.4	27.1	724	12	CNS05DSJ
C 25	41.4	27.1	734	12	CNS056MU
C 26	41.4	27.1	923	12	CNS04J5I
C 27	41.4	27.1	963	12	CNS01X68
C 28	41.4	27.1	1015	12	CNS05KUD
C 29	41.4	27.1	1047	12	CNS04JYJ
C 30	41.4	27.1	1050	12	CNS05M3J
C 31	41.4	27.1	1054	12	CNS0506W
C 32	41.4	27.1	1074	12	CNS05IOP
C 33	41.4	27.1	1096	12	CNS05IC3
C 34	41	26.8	812	12	CNS036TC
C 35	40.2	26.3	1017	12	CNS033G4
C 36	40.2	26.3	1122	12	CNS05F9J
C 37	39.8	26.0	297	10	H53538
C 38	39.8	26.0	967	12	CNS02BFS
C 39	39.8	26.0	1029	12	CNS05GSM
C 40	39.8	26.0	1036	12	CNS04VNY
C 41	39.8	26.0	1038	12	CNS05AVS
C 42	39.8	26.0	1051	12	CNS02H6W
C 43	39.8	26.0	1056	12	CNS05R9Z
C 44	39.8	26.0	1079	12	CNS05HIG
C 45	39.8	26.0	1080	12	CNS054WU

## ALIGNMENTS

### RESULT 1

LOCUS BB155356  
DEFINITION BB155356 RIKEN full-length enriched, 16 days neonate thymus Mus musculus cDNA clone AL30024116 3' similar to L10410 Mouse DNA-binding protein (CHD-1) mRNA, mRNA sequence.  
ACCESSION BB155356  
VERSION BB155356.2  
KEYWORDS GI:16268254  
SOURCE EST.  
ORGANISM house mouse.  
REFERENCE 1 Mus musculus  
AUTHORS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus. (Bases 1 to 619)  
Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.  
RIKEN Mouse ESTs (Arakawa,T., et al. 2001)  
Unpublished (2001)  
On Jun 29, 2000 this sequence version replaced gi:8811286.  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@gsr.riken.go.jp,  
URL:http://genome.gsc.riken.go.jp/  
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)  
wagi,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura



prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGATTCTCGAGTTAAATAAATATCCCCCCCCCC 3']. cDNA was cleaved with BamHI and XhoI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda PLC I."

BASE COUNT 211 a 130 c 168 g 151 t  
ORIGIN

Query Match 45.1%; Score 69; DB 9; Length 660;  
Best Local Similarity 80.2%; Pred. No. 5.9e-10; Indels 0; Gaps 0;  
Matches 81; Conservative 0; Mismatches 20;

QY 52 AAGAACCCCGAGCTAGCAGTTACAGACCCGTCGAGATTACCTCATTAATAATTACTGAAT 111  
Db 502 AAAAAACCAAGCAAGCAACAGTTACAGACCCGTCGAGACTCATCAAACTACTTAGC 561

QY 112 AAGACCTTGCAAGAAAGAACGACAGACTTGTGTGTC 152

Db 562 AGAGATCTTGCAAAAAGAGAGGCTCAGAGACTTGTGTGTC 602

## RESULT 3

AL659353/c  
LOCUS AL659353 593 bp mRNA linear EST 13-DEC-2001  
DEFINITION AL659353 XGC-neurula silurana tropicalis cDNA clone TNeu045e20 5', mRNA sequence.

ACCESSION AL659353

VERSION AL659353.1 GI:17672995

KEYWORDS EST.

SOURCE western clawed frog.

## ORGANISM

Silurana tropicalis

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;

Xenopodinae; Silurana.

1 (bases 1 to 593)

Huckle, E., Taylor, R., Ashurst, J.L., Zorn, A.M. and Rogers, J.

Sanger Xenopus tropicalis EST project 2001 (10\_2001)

Unpublished (2001)

Contact: Huckle E

Sanger Centre

Hinxton, Cambridgeshire, CB10 1SA, UK

Email: trop@sanger.ac.uk

Sanger Xenopus tropicalis EST project 2001

TROPICALIS\_SEQUENCE\_ID: TNeu045e20.sp6

Sequencing primer: SP6

This sequence is from a Xenopus Gene Collection (XGC) library

constructed by Aaron M. Zorn.

Location/Qualifiers

1..593

/organism="Silurana tropicalis"

/db\_xref="taxon:8364"

/clone="TNeu045e20"

/clone\_lib="XGC-neurula"

/dev\_stage="neurula"

/lab\_host="Escherichia coli DH10B"

/note="Vector: PCS107; Site\_1: EcoRI; Site\_2: NotI; cDNA

was oligo dT primed from 5' end of poly A+ RNA from neurula.

EcoRI-NotI cut cDNA was then ligated into PCS107 with

EcoRI at the 5' end and NotI at the 3' end."

## BASE COUNT

119 a 134 c 114 g 226 t

## ORIGIN

Query Match 43.7%; Score 66.8; DB 9; Length 593;  
Best Local Similarity 78.4%; Pred. No. 2.5e-09;  
Matches 80; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

QY 52 AAGAACCCCGAGCTAGCAGTTACAGACCCGTCGAGATTACCTCATTAATAATTACTGAAT 111  
Db 325 AAGAACCCCGAGCTAGCAGTTACAGACCCGTCGAGACTCATTAATAATTACTGAAT 266

QY 112 AAGACCTTGCAAGAAAGAACGACAGACTTGTGTGTC 153  
Db 265 AAGATCTGCTTAGGAAAGAACGACAAAGACTTCTAGTGGA 224

## RESULT 4

AL644594  
LOCUS AL644594 645 bp mRNA linear EST 07-NOV-2001  
DEFINITION AL644594 XGC-egg silurana tropicalis cDNA clone L1E1d12 5', mRNA sequence.

ACCESSION AL644594

VERSION AL644594.1 GI:16796719

KEYWORDS EST.

SOURCE western clawed frog.

## ORGANISM

Silurana tropicalis

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;

Xenopodinae; Silurana.

1 (bases 1 to 645)

Huckle, E., Taylor, R., Ashurst, J.L., Zorn, A.M. and Rogers, J.

Sanger Xenopus tropicalis EST project 2001 (10\_2001)

Unpublished (2001)

Contact: Huckle E

Sanger Centre

Hinxton, Cambridgeshire, CB10 1SA, UK

Email: trop@sanger.ac.uk

Sanger Xenopus tropicalis EST project 2001

TROPICALIS\_SEQUENCE\_ID: L1E1d12.plc

Sequencing primer: P1C

This sequence is from a Xenopus Gene Collection (XGC) library

constructed by Aaron M. Zorn.

Location/Qualifiers

1..645

/organism="Silurana tropicalis"

/db\_xref="taxon:8364"

/clone="L1E1d12"

/clone\_lib="XGC-egg"

/dev\_stage="egg"

/lab\_host="Escherichia coli XL1-blue"

/note="Vector: PCS107; Site\_1: EcoRI; Site\_2: NotI; cDNA

was oligo dT primed from 5' end of poly A+ RNA from egg.

EcoRI-NotI cut cDNA was then ligated into PCS107 with

EcoRI at the 5' end and NotI at the 3' end."

222 a 125 c 156 g 141 t 1 others

BASE COUNT 222 a 125 c 156 g 141 t 1 others

ORIGIN

Query Match 43.7%; Score 66.8; DB 9; Length 645;

Best Local Similarity 78.4%; Pred. No. 2.6e-09;

Matches 80; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

QY 52 AAGAACCCCGAGCTAGCAGTTACAGACCCGTCGAGATTACCTCATTAATAATTACTGAAT 111

Db 477 AAGAACCCCGAGCTAGCAGTTACAGACCCGTCGAGATTACCTCATTAATAATTACTGAAT 536

QY 112 AAGACCTTGCAAGAAAGAACGACAGACTTGTGTGTC 153

Db 537 AAGATCTGCTTAGGAAAGAACGACAAAGACTTCTAGTGGA 578

## RESULT 5

BB834922  
LOCUS BB834922 446 bp mRNA linear EST 19-NOV-2001  
DEFINITION BB834922 RIKEN full-length enriched, mammary gland RCB-0527  
JYG-MC(B) cDNA Mus musculus cDNA clone G930033J21 3', mRNA sequence.

ACCESSION BB834922

VERSION BB834922.1 GI:17013165

KEYWORDS EST.

SOURCE house mouse.

## ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;







QY	111	TAAAGACCTTGCAAGAAAGGACGACAGAGACTTGCTGGTGC	152
Db	381	CAGAGACTCTGCAAAAGACAGAGCTCAGAGACTTGCTGGTGC	422
RESULT	12		
LOCUS	AW997058/c		
DEFINITION	QV3-BN0047-150400-152-c03 BN0047 Homo sapiens cDNA, mRNA sequence.		
ACCESSION	AW997058		
VERSION	AW997058.1	GI:8257292	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE			
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 686)		
TITLE	Dias Neto E., Garcia Correa R., Verjovsky-Almeida S., Briones M.R., Nagai M.A., da Silva W. Jr., Zago M.A., Bordin S., Costa F.F., Goldman G.H., Carvalho A.F., Matsukuma A., Baia G.S., Simpson D.H., Brunstein A., de Oliveira P.S., Bucher P., Jongeneel C.V., O'Hare M.J., Soares F., Brentani R.R., Reis L.F., de Souza S.J. and Simpson A.J.		
JOURNAL	Shotgun sequencing of the human transcriptome with ORF expressed sequence tags		
COMMENT	Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000) 2020263 Contact: Simpson A.J.G. Laboratory of Cancer Genetics Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil Tel: +55-11-2704922 Fax: +55-11-2707001 Email: asimpson@ludwig.org.br This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=et3-QV3-BN0047-150400-152-c03&t3=2000-04-15&t4=1) Seq primer: puc 18 forward High quality sequence start: 19 High quality sequence stop: 678.		
FEATURES			
SOURCE	1..686 /organism="Homo sapiens" /db_xref="taxon:9606" /clone_lib="BN0047" /dev_stage="Adult" /note="Organ: breast_normal; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."		
BASE COUNT	165 a 154 c 126 g 241 t		
ORIGIN			
Query Match	35.4%;	Score 54.2;	DB 9; Length 686;
Best Local Similarity	82.7%;	Pred. No. 1.2e-05;	
Matches	62;	Conservative 0;	Mismatches 13; Indels 0; Gaps 0;
QY	53	AGAAACCCAGGCTAAGCAGTTACAGACCGTGCAGATTACCTTCATTAAATTAATCTGAATA	112
Db	106	AAAAACCAAGCAAAACAGTTGCAGACCCGTGCAGACTACCTCATCAATTAATCTAGTA	47
QY	113	RAGACCTTCAGAA	127
Db	46	GAGATCTTGCAAAA	32
RESULT	13		
LOCUS	CNS04DVG/c		
DEFINITION	Tetraodon nigroviridis genome survey sequence T7 end of clone 103K08 of library G from Tetraodon nigroviridis, genomic survey sequence.		
ACCESSION	AL286261		
VERSION	AL286261.1	GI:8024707	
KEYWORDS	GSS; genome survey sequence.		
SOURCE	Tetraodon nigroviridis		
ORGANISM	Tetraodon nigroviridis		
REFERENCE			
AUTHORS	Bounau L., Billault A., Quetier F., Saurin W., Bernot A. and Weissenbach J.		
TITLE	Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 856)		
AUTHORS	Roest-Crollius H., Jaillon O., Dasilva C., Bouneau L., Fisher C., Bernot A., Fizames C., Wincker P., Brottier P., Quetier F., Saurin W. and Weissenbach J.		
TITLE	Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence		
JOURNAL	Unpublished		
REFERENCE	3 (bases 1 to 856)		
AUTHORS	Direct Submission		
JOURNAL	Submitted (12-APR-2000) to the EMBL/Genbank/DBJ databases		
COMMENT	This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/Tetraodon.		
FEATURES			
SOURCE	1..856 /organism="Tetraodon nigroviridis" /db_xref="taxon:99883" /clone="103K08" /clone_lib="G" /note="Genoscope sequence ID : CORG103BF04LP1-end : T7"		
BASE COUNT	84 a 308 c 252 g 206 t		
ORIGIN			
Query Match	34.8%;	Score 53.2;	DB 12; Length 856;
Best Local Similarity	69.4%;	Pred. No. 2.5e-05;	
Matches	68;	Conservative 3;	Mismatches 27; Indels 0; Gaps 0;
QY	52	AAGAAACCCAGGCTAAGCAGTTACAGACCGTGCAGATTACCTTCATTAAATTAATCTGAAT	111
Db	656	AAGAAGCTCAGCCAAACAGTTACAGACCGGACGAGCCDACCCTACCTCATCARGCTGTGAGC	597
QY	112	AAAGACCTTGCAGAAAGCAAGACACAGACTTGCTGG	149
Db	596	AAGCAGCTGGCCAGAAAGAACCCCAAGCAGGSGGG	559
RESULT	14		
LOCUS	BI844947/c		
DEFINITION	ql17e05 x1 zebrafish adult brain Danio rerio cDNA clone 4831593 3' similar to SW-CHD2_HUMAN O14647 CHROMODOMAIN-HELICASE-DNA-BINDING PROTEIN 2 ; mRNA sequence.		
ACCESSION	BI844947		
VERSION	BI844947.1	GI:15957470	
KEYWORDS	EST.		
SOURCE	zebrafish.		
ORGANISM	Danio rerio		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		

Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes  
 ; Cyprinidae; Danio.  
 1 (bases 1 to 532)  
 Clark, M., Johnson, S.L., Lehrach, H., Lee, R., Li, F., Marra, M., Eddy  
 , S., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wyllie, T., Underwood  
 , K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B.,  
 Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E.,  
 Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R.  
 and Wilson, R.

TITLE  
 JOURNAL  
 COMMENT

WashU zebrafish EST Project 1998  
 Unpublished (1998)  
 Other\_ESTs: fql7e05.y1  
 Contact: Stephen L. Johnson  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810

Email: zbrafish@watson.wustl.edu  
 CDNA Library Preparation: John Ngai. CDNA Library Arrayed by:  
 Matthew Clark. DNA Sequencing by: Washington University Genome  
 Sequencing Center Clone Distribution: Genome Systems, St. Louis,  
 Missouri (web address: www.genomesystems.com) (email contact:  
 info@genomesystems.com) and Research Genetics, Huntsville, Alabama  
 (web address: www.resgen.com) (email contact: info@resgen.com) and  
 Ressourcenzentrum Primardatenbank, Berlin, Germany (web address:  
 www.rzpd.de)

Trace considered overall poor quality

Seq primer: T7 from Gibco

High quality sequence stop: 1.

FEATURES  
 source

Location/Qualifiers

1..532  
 /organism="Danio rerio"  
 /db\_xref="taxon:7955"  
 /clone="4831593"  
 /clone\_lib="zebrafish adult brain"  
 /sex="mixed male and female"  
 /tissue\_type="brain"  
 /dev\_stage="adult"  
 /lab\_host="E. coli DH10B"  
 /note="Vector: pZIPLOX; Site\_1: NotI; Site\_2: SalI;  
 Original library was constructed in lambdaZiPLOX. Mass  
 excision of the cDNA library was performed to yield  
 pZIPLOX plasmids. Insert check was done in original  
 library."

BASE COUNT 88 a 128 c 104 g 212 t  
 ORIGIN

Query Match 30.2%; Score 46.2; DB 10; Length 532;  
 Best Local Similarity 72.3%; Pred. No. 0.0025;  
 Matches 60; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 52 AAGAACCCAGGCTAAGCAGTTACAGACCCGTCGAGATTACCTCATTAATTAATCTGAAT 111

DB 349 AAGAACCCTCAAGGAAGCAGTTACAGATGCGAACAGACTACTTGTCTGAAGATGCTGAAG 290

QY 112 AAAGACCTTGCAGGAAGGAAGC 134

DB 289 AAGACCTGGAAGCAAGCAAGATGC 267

RESULT 15  
 BM185888/c  
 LOCUS BM185888 574 bp mRNA linear EST 11-DEC-2001  
 DEFINITION fV74a09.x1 zebrafish adult brain Danio rerio cDNA clone 5413937 3'  
 similar to SW:CHD2\_HUMAN O14647 CHROMODOMAIN-HELICASE-DNA-BINDING  
 PROTEIN 2 ; mRNA sequence.  
 ACCESSION BM185888  
 VERSION BM185888.1 GI:17516846  
 KEYWORDS EST.  
 SOURCE zebrafish.  
 ORGANISM Danio rerio.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes  
 ; Cyprinidae; Danio.  
 1 (bases 1 to 574)

REFERENCE  
 AUTHORS

Clark, M., Johnson, S.L., Lehrach, H., Lee, R., Li, F., Marra, M., Eddy  
 , S., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wyllie, T., Underwood  
 , K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B.,  
 Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E.,  
 Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R.  
 and Wilson, R.

TITLE  
 JOURNAL  
 COMMENT

WashU zebrafish EST Project 1998  
 Unpublished (1998)  
 Other\_ESTs: fv74a09.y1  
 Contact: Stephen L. Johnson  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810

Email: zbrafish@watson.wustl.edu  
 CDNA Library Preparation: John Ngai. CDNA Library Arrayed by:  
 Matthew Clark. DNA Sequencing by: Washington University Genome  
 Sequencing Center Clone Distribution: Genome Systems, St. Louis,  
 Missouri (web address: www.genomesystems.com) (email contact:  
 info@genomesystems.com) and Research Genetics, Huntsville, Alabama  
 (web address: www.resgen.com) (email contact: info@resgen.com) and  
 Ressourcenzentrum Primardatenbank, Berlin, Germany (web address:  
 www.rzpd.de)

Seq primer: T7 from Gibco

High quality sequence stop: 328.

FEATURES  
 source

Location/Qualifiers

1..574  
 /organism="Danio rerio"  
 /db\_xref="taxon:7955"  
 /clone="5413937"  
 /clone\_lib="zebrafish adult brain"  
 /sex="mixed male and female"  
 /tissue\_type="brain"  
 /dev\_stage="adult"  
 /lab\_host="E. coli DH10B"  
 /note="Vector: pZIPLOX; Site\_1: NotI; Site\_2: SalI;  
 Original library was constructed in lambdaZiPLOX. Mass  
 excision of the cDNA library was performed to yield  
 pZIPLOX plasmids. Insert check was done in original  
 library."

BASE COUNT 100 a 135 c 116 g 223 t  
 ORIGIN

Query Match 30.2%; Score 46.2; DB 10; Length 574;  
 Best Local Similarity 72.3%; Pred. No. 0.0026;  
 Matches 60; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 52 AAGAACCCAGGCTAAGCAGTTACAGACCCGTCGAGATTACCTCATTAATTAATCTGAAT 111

DB 349 AAGAACCCTCAAGGAAGCAGTTACAGATGCGAACAGACTACTTGTCTGAAGATGCTGAAG 290

QY 112 AAAGACCTTGCAGGAAGGAAGC 134

DB 289 AAGACCTGGAAGCAAGCAAGATGC 267

Search completed: August 2, 2002, 16:52:15  
 Job time: 12216 sec



Mon Aug 5 11:51:53 2002

us-08-973-363-4.rst

Page 9

**THIS PAGE BLANK (USPTO)**



PI Griffiths R, Tiwari B;  
 XX WPI; 1997-043127/04.  
 DR P-PSDB; AAW08149.  
 XX Avian chromodomain-helicase-DNA binding genes determine sex in  
 PT birds - used for sex determin. and to control sex of progeny  
 XX  
 PS Claim 8; Fig 3; 76pp; English.  
 XX  
 CC Bases 3855-3977 (AAT42756) of the mouse CHD-1 gene show homology  
 CC to portions of the chicken CHD-1A (A = Avian) gene (AAT42757),  
 CC chicken CHD-W (W refers to the W chromosome) gene (AAT42758) and  
 CC and the great tit CHD-W gene (AAT42759). Translated amino acid  
 CC sequences of this region are provided in AAW08146-49. The CHD-1A  
 CC (see also AAT42751) and CHD-W (see also AAT42754-55) genes determine  
 CC sex in birds and can be used to identify the sex of an embryo,  
 CC foetus etc. and to manipulate the sex of progeny.  
 XX  
 SQ Sequence 153 BP; 60 A; 33 C; 31 G; 29 T; 0 other;

Query Match 99.0%; Score 151.4; DB 18; Length 153;  
 Best Local Similarity 99.3%; Pred. No. 4e-36;  
 Matches 152; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATTTTACCTGATGACCCAGATAGAAACACAGGCAAGCAGTTGCAGACCAAGAAACCA 60  
 ||||||||||| ||||||||||| ||||||||||| ||||||||||| |||||||||||  
 Db 1 attttacctgatgccagataagaacacagcaggcaagcagcttcagaccagaagaaacca 60  
 ||||||||||| ||||||||||| ||||||||||| ||||||||||| |||||||||||

QY 61 CAGGCAAGCAGTTGCAGACCCGTCAGATTACCTCATTAATTACTGAATAAAGACCTT 120  
 ||||||||||| ||||||||||| ||||||||||| ||||||||||| |||||||||||  
 Db 61 caggcaagcagttgcagaccctgcagattacctcattaaattactgaataaagacctt 120  
 ||||||||||| ||||||||||| ||||||||||| ||||||||||| |||||||||||

QY 121 GCAAGAAAGAGTGCAAGACTTACTGTGCA 153  
 ||||||||||| ||||||||||| ||||||||||| ||||||||||| |||||||||||  
 Db 121 gcaagaaagagtgcaagacttactgtgtgca 153

RESULT 2  
 AAT42758  
 ID AAT42758 standard; DNA; 153 BP.  
 XX  
 AC AAT42758;  
 XX  
 DT 12-MAR-1997 (first entry)  
 XX  
 DE Chick CHD-W gene fragment.  
 XX  
 KW Bird; sex determination; chromodomain-Helicase-DNA binding 1;  
 KW CHD-1A; CHD-W; W chromosome; ss.  
 XX  
 OS Gallus sp.  
 XX  
 FH Key Location/Qualifiers  
 FT misc\_difference 52..81  
 FT /\*tag= a  
 FT /note= "bases 52-81 are a repeat of bases 22-51  
 FT and are ignored in the translated amino  
 FT acid sequence given in Fig 3"  
 XX  
 PN W09639505-A1.  
 XX  
 PD 12-DEC-1996.  
 XX  
 PF 05-JUN-1996; 96WO-GB01341.  
 XX  
 PR 06-JUN-1995; 95GB-0011439.  
 XX  
 PA (ISIS-) ISIS INNOVATION LTD.  
 XX  
 PI Griffiths R, Tiwari B;  
 XX  
 DR WPI; 1997-043127/04.  
 DR P-PSDB; AAW08147.

DR WPI; 1997-043127/04.  
 DR P-PSDB; AAW08148.  
 XX  
 PT Avian chromodomain-helicase-DNA binding genes determine sex in  
 PT birds - used for sex determin. and to control sex of progeny  
 XX  
 PS Claim 8; Fig 3; 76pp; English.  
 XX  
 CC Bases 3855-3977 (AAT42756) of the mouse CHD-1 gene show homology  
 CC to portions of the chicken CHD-1A (A = Avian) gene (AAT42757),  
 CC chicken CHD-W (W refers to the W chromosome) gene (AAT42758) and  
 CC and the great tit CHD-W gene (AAT42759). Translated amino acid  
 CC sequences of this region are provided in AAW08146-49. The CHD-1A  
 CC (see also AAT42751) and CHD-W (see also AAT42754-55) genes determine  
 CC sex in birds and can be used to identify the sex of an embryo,  
 CC foetus etc. and to manipulate the sex of progeny.  
 XX  
 SQ Sequence 153 BP; 56 A; 36 C; 31 G; 30 T; 0 other;

Query Match 87.5%; Score 133.8; DB 18; Length 153;  
 Best Local Similarity 92.2%; Pred. No. 7.5e-31;  
 Matches 141; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 1 ATTTTACCTGATGACCCAGATAGAAACACAGGCAAGCAGTTGCAGACCAAGAAACCA 60  
 ||||||||||| ||||||||||| ||||||||||| ||||||||||| |||||||||||  
 Db 1 attttacctgatgccagataagaacacccagcaggcttaagcagaccagaagaaaccc 60  
 ||||||||||| ||||||||||| ||||||||||| ||||||||||| |||||||||||

QY 61 CAGGCAAGCAGTTGCAGACCCGTCAGATTACCTCATTAATTACTGAATAAAGACCTT 120  
 ||||||||||| ||||||||||| ||||||||||| ||||||||||| |||||||||||  
 Db 61 caggcaagcagttgcagaccctgcagattacctcattaaattactgaataaagacctt 120  
 ||||||||||| ||||||||||| ||||||||||| ||||||||||| |||||||||||

QY 121 GCAAGAAAGAGTGCAAGACTTACTGTGCA 153  
 ||||||||||| ||||||||||| ||||||||||| ||||||||||| |||||||||||  
 Db 121 gcaagaaagagtgcaagacttactgtgtgca 153

RESULT 3  
 AAT42757  
 ID AAT42757 standard; DNA; 153 BP.  
 XX  
 AC AAT42757;  
 XX  
 DT 12-MAR-1997 (first entry)  
 XX  
 DE Chick CHD-1A gene fragment.  
 XX  
 KW Bird; sex determination; chromodomain-Helicase-DNA binding 1;  
 KW CHD-1A; CHD-W; W chromosome; ss.  
 XX  
 OS Gallus sp.  
 XX  
 FH Key Location/Qualifiers  
 FT misc\_difference 52..81  
 FT /\*tag= a  
 FT /note= "bases 52-81 are a repeat of bases 22-51  
 FT and are ignored in the translated amino  
 FT acid sequence given in Fig 3"  
 XX  
 PN W09639505-A1.  
 XX  
 PD 12-DEC-1996.  
 XX  
 PF 05-JUN-1996; 96WO-GB01341.  
 XX  
 PR 06-JUN-1995; 95GB-0011439.  
 XX  
 PA (ISIS-) ISIS INNOVATION LTD.  
 XX  
 PI Griffiths R, Tiwari B;  
 XX  
 DR WPI; 1997-043127/04.  
 DR P-PSDB; AAW08147.

birds - used for sex determin. and to control sex of progeny  
 Claim 8; Fig 3; 76pp; English.  
 Bases 3855-3977 (AAT42756) of the mouse CHD-1 gene show homology to portions of the chicken CHD-1A (A = Avian) gene (see also AAT42757), chicken CHD-W (W refers to the W chromosome) gene (see also AAT42758) and the great tit CHD-W gene (see also AAT42759). Translated amino acid sequences of this region are provided in AAW08146-49. The CHD-1A (see also AAT42751) and CHD-W (see also AAT42754-55) genes determine sex in birds and can be used to identify the sex of an embryo, foetus etc. and to manipulate the sex of progeny.  
 Sequence 153 BP: 61 A; 37 C; 27 G; 28 T; 0 other;  
 SO

	Query Match	69.0%;	Score 105.6;	DB 18;	Length 153;
	Best Local Similarity	80.9%;	Prod. No. 2.1e-22;		
	Matches 123;	Conservative	0;	Mismatches 29;	Indels 0;
	Gaps	0;			
QY	1	ATTTTACTGTATGACCAGATAAAGAACACACAGGCAAAAGCTGTGCAGACCAAGAAACCA	60		
Db	1	attcttcagatgatcctgataaaaaaccacaagcaaacaggtttacagaccacaaacacca	60		
QY	61	CAGGCAAAAGCTGTGCAGACCGCTGCAGATTAACCTCATTAATTAATTAAGAAAGCTTT	120		
Db	61	caagcaaacaggtttacagaccgctgcagactacctcaactacttagcagagatcctt	120		
QY	121	GCAAGAAAGAAAGTGCAAAGACTTACTTGGTGC	152		
Db	121	qcaaaaadagagqctcaagagactttgqgtgc	152		

RESULT	5	
AAT42751		
ID	AAT42751	standard; cDNA; 5608 BP.
XX		
AC	AAT42751;	
DT	12-MAR-1997	(first entry)
DE	Chicken CHD-1A gene.	
DE		
XX		
KW	Bird; sex determination; chromodomain-Helicase-DNA binding 1 Avian;	
KW	CHD-W; chromodomain-Helicase-DNA binding on the W chromosome; ss.	
XX		
OS	Gallus sp.	
XX		
FH	Key	Location/Qualifiers
FT	CDS	228..5390
FT		/*tag= a
XX		
PN	W09639505-A1.	
XX		
PD	12-DEC-1996.	
XX		
PF	05-JUN-1996;	96WO-GB01341.
XX		
PR	06-JUN-1995;	95GB-0011439.
XX		
PA	(ISIS-) ISIS INNOVATION LTD.	
XX		
PI	Griffiths R, Tiwari B;	
XX		
WPI	1997-043127/04.	
XX		
PT	Avian chromodomain-helicase-DNA binding genes determine sex in	
PT	birds - used for sex determ. and to control sex of progeny	
XX		
PS	Claim 1: Fig 5; 76pp; English.	
XX		
CC	The chicken CHD-W gene (AAT42754) acting alone or in conjunction with	

CC the closely related CHD-1A gene (AAT42751) is suggested to initiate  
CC female development in birds. The sequence of CHD-1A was deduced  
CC from 3 clones isolated from a stage 10-12 chicken embryo cDNA  
CC library using a great tit CHD-W sequence (see also AAT42755) as probe.  
CC The CHD-1A (A = Avian) gene shows close identity to the mouse CHD-1  
CC gene (see also AAT42756-57). It is located on an autosome or Z  
CC chromosome. Probes based on CHD-W and CHD-1A give a W chromosome-  
CC specific signal on hybridisation to genomic DNA of a non-rare  
CC bird and can be used for sex determin. of a bird. CHD-1A nucleic  
CC acids can also be used to control the sex of the progeny of a bird.  
XX  
SQ Sequence 6608 BP; 2289 A; 1207 C; 1459 G; 1653 T; 0 other;

Query Match 57.9%; Score 88.6; DB 18; Length 6608;  
Best Local Similarity 91.3%; Pred. No. 7.8e-17;  
Matches 94; Conservative 0; Mismatches 9; Indels 0; Gaps 0;  
QY 51 CAAGAAACCACGAGCAAGCAGTTCGACGCCGTCGAGATTACCTCATTAAATTACTGAA 110  
D 110  
Db 4100 caagaaccacccagcgaagcagctacagaccgctgcagactacccattaaattactgaa 4159  
QY 111 TAAAGACCTTGCAGAAAGAGAGTGCAGAGACTTACTGGTGCA 153  
D 153  
Db 4160 taaagaccttgcaagaaggaagcacaagggctgtggtgca 4202

RESULT 6  
AAT42754  
ID AAT42754 standard; cDNA; 1316 BP.  
XX  
AC AAT42754;  
XX  
DT 12-MAR-1997 (first entry)  
XX  
DE Chicken CHD-W gene (partial sequence).  
XX

XX Bird; sex determination; chromodomain-Helicase-DNA binding 1 Avian;  
KW CHD-W; chromodomain-Helicase-DNA binding on the W chromosome; ss.  
XX  
XX Gallus sp.  
XX  
XX WO9639505-Al.  
XX  
XX 12-DEC-1996.  
XX  
XX 05-JUN-1996; 96WO-GB01341.  
XX  
XX 06-JUN-1995; 95GB-0011439.  
XX  
XX (ISIS-) ISTIS INNOVATION LTD.  
XX  
XX Griffiths R, Tiwari B;  
XX  
XX WPI; 1997-043127/04.  
XX  
XX Avian chromodomain-helicase-DNA binding genes determine sex in  
PT birds - used for sex determin. and to control sex of progeny  
XX  
XX Claim 1; Fig 8; 76pp; English.

XX The chicken CHD-W gene (AAT42754) acting alone or in conjunction with  
CC the closely related CHD-1A gene (AAT42751) is suggested to initiate  
CC female development in birds. The sequence of CHD-W was deduced  
CC from 2 clones isolated from a 10-day chick embryo library using  
CC a fragment of the CHD-1A gene as a probe. The CHD-W gene is  
CC located on the W chromosome. Probes based on CHD-W and CHD-1A give  
CC a W chromosome-specific signal on hybridisation to genomic DNA of a  
CC non-rare bird and can be used for sex determin. of a bird. CHD-W  
CC nucleic acids can also be used to control the sex of progeny of a  
CC bird.  
XX

SQ Sequence 1316 BP; 492 A; 208 C; 306 G; 304 T; 6 other;

Query Match 57.3%; Score 87.6; DB 18; Length 1316;  
Best Local Similarity 91.2%; Pred. No. 9.8e-17;  
Matches 93; Conservative 0; Mismatches 9; Indels 0; Gaps 0;  
QY 52 AAGAAACACAGGCAAGCAGTTCGACGCCGTCGAGATTACCTCATTAAATTACTGAAT 111  
D 111  
Db 1159 aagaaccacccagcgaagcagctacagactcgtgcagattacccattaaattactgaat 1218  
QY 112 AAGACCTTGCAGAAAGAGAGTGCAGAGACTTACTGGTGCA 153  
D 153  
Db 1219 aagaaccttgcaagaaggaagcacaagggctgtggtgca 1260

RESULT 7  
AAV59280  
ID AAV59280 standard; cDNA; 1311 BP.  
XX  
AC AAV59280;  
XX  
DT 14-DEC-1998 (first entry)  
XX  
DE Altered telomere repeat binding factor 1 gene.  
XX  
KW ds; human; telomere repeat binding factor; A-TRF; dimerisation domain;  
KW telomere; ageing; ataxia telangiectasia; Down's syndrome; tumour; viral.  
XX  
XX Homo sapiens.  
OS Synthetic.  
XX  
FH Key Location/Qualifiers  
FT CDS 1..1311  
FT FT /\*tag= a  
FT FT /product= "A-TRF"  
XX  
XX WO9836066-Al.  
XX  
XX 20-AUG-1998.  
XX  
XX 13-FEB-1998; 98WO-US02765.  
XX  
XX 04-FEB-1998; 98US-0018628.  
XX  
XX 13-FEB-1997; 97US-0800264.  
XX  
XX (UVRQ ) UNIV ROCKEFELLER.  
XX  
XX Bianchi A, De Lange T, Van Steensel B;  
XX  
XX WPI; 1998-480769/41.  
XX  
XX P-PSDB; AAV59280.

XX Nucleic acid encoding altered telomere repeat binding protein and  
PT related vectors - transformants, hetero-dimers and antibodies, used  
PT to inhibit shortening of telomerases caused by ageing or disease,  
PT also used to extend life of cells in culture  
XX

XX Claim 14; Page 110-111; 163pp; English.

XX The altered vertebrate telomere repeat binding protein (A-TRF) has a  
CC telomere repeat binding factor (TRF) dimerisation domain, and forms a  
CC hetero-dimer with TRF, preventing it from binding to the specified repeat  
CC sequence. A-TRF, optionally expressed by gene therapy, is used to  
CC inhibit shortening of telomerases associated with ageing (for cosmetic  
CC purposes) and disease, e.g. ataxia telangiectasia, Down's syndrome,  
CC atrophy of the skin, age-related macular degeneration, atherosclerosis,  
CC tumours and viral (including human immune deficiency virus) infection.  
CC Cells expressing A-TRF also have an increased life span in vitro, e.g.  
CC for expression of recombinant proteins or where intended for subsequent  
CC transplant or for testing, eliminating the need for transformation.  
XX

SQ Sequence 1311 BP; 441 A; 257 C; 316 G; 297 T; 0 other;

Query Match	28.0%	Score 42.8;	DB 19;	Length 1311;
Best Local Similarity	87.0%;	Pred. NO. 0.0026;		
Matches 47;	Conservative 0;	Mismatches 7;	Indels 0;	Gaps 0;
QY	52	AAGAAACACAGGCAAGCAGTGTGCAGACCCGTGCAGATTACCTCATTAAATTA	105	
Db	1258	aaaaaacacacagcaaacagttacacccgtgacactacotcatcaacta	1311	
RESULT	8			
ABL06443				
ID	ABL06443	standard; cDNA; 6240 BP.		
XX	ABL06443;			
XX				
DT	26-MAR-2002	(first entry)		
DE				
XX	Drosophila melanogaster	expressed polynucleotide SEQ ID NO 13811.		
KW	Drosophila;	developmental biology; cell signalling; insecticide;		
KW	pharmaceutical; gene; ss.			
XX				
XX	Drosophila melanogaster.			
OS				
PN	WO200171042-A2.			
PD				
PD	27-SEP-2001.			
XX				
PF	23-MAR-2001;	2001WO-US09231.		
XX				
PR	23-MAR-2000;	2000US-191637P.		
PR	11-JUL-2000;	2000US-0614150.		
XX				
PA	(PEKE )	PE CORP NY.		
PI				
PI	Venter JC, Adams M, Li PWD, Myers EW;			
XX				
XX	WPI: 2001-658860/75.			
XX	P-PSDB; ABB62340.			
XX				
PT	New isolated nucleic acid detection reagent for detecting 1000 or more			
PT	genes from Drosophila and for elucidating cell signalling and cell-cell			
PT	interactions -			
PS	Claim 1; SEQ ID NO 13811; 21pp + Sequence Listing; English.			
CC	The invention relates to an isolated nucleic acid detection reagent			
CC	capable of detecting 1000 or more genes from Drosophila. The invention is			
CC	useful in developmental biology and in elucidating cell signalling and			
CC	cell-cell interactions in higher eukaryotes for the development of			
CC	insecticides, therapeutics and pharmaceutical drugs. The invention of			
CC	discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA			
CC	sequences (ABL01840-ABL16175) and the encoded proteins			
CC	(ABB57737-ABB72072).			
CC	The sequence data for this patent did not form part of the printed			
CC	specification, but was obtained in electronic format directly from WIPO			
CC	at ftp.wipo.int/pub/published_pct_sequences.			
XX	Sequence 6240 BP; 1675 A; 1682 C; 1760 G; 1123 T; 0 other:			

[illegible]

DT 05-NOV-2001 (first entry)  
XX Human digestive system antigen coding sequence SEQ ID NO: 1198.  
DE Human digestive system antigen; gene therapy; cancer; appendicitis;  
XX ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;  
KW digestive system disorder; Meckel's diverticulum; ss.  
XX Homo sapiens.  
XX WO20015314-A2.  
PN 02-AUG-2001.  
PD 17-JAN-2001; 2001WO-US01324.  
XX 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235835.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 06-DEC-2000; 2000US-0251856.



RPR	08-DEC-2000; 2000US-0251868.
RPR	08-DEC-2000; 2000US-0251869.
RPR	08-DEC-2000; 2000US-0251889.
RPR	08-DEC-2000; 2000US-0251989.
RPR	11-DEC-2000; 2000US-0251990.
RPR	11-DEC-2000; 2000US-0254097.
RPR	05-JAN-2001; 2001US-0259678.
XX	(HUMA-) HUMAN GENOME SCI INC.
RPA	Rosen CA, Barash SC, Ruben SM;
XX	WPI; 2001-503630/55.
PFI	P-PSDB; AAM93109.
XXX	
DNR	Polynucleotides encoding digestive system antigens, useful for
PTT	diagnosing, treating, preventing and/or prognosing disorders of the
PTP	digestive system, particularly cancer and cancer metastases -
XXX	
PS	Claim 1; SEQ ID NO 1198; 986pp; English.
CCC	The present invention provides the protein and coding sequences of a
CCC	number of human digestive system antigens. These can be used in the
CCC	diagnosis, treatment and prevention of digestive system disorders,
CCC	including cancer, Meckel's diverticulum, bacterial or parasitic
CCC	Infection, appendicitis, Hirschsprung's disease, chronic colitis or
CCC	ulcerative colitis. The present sequence is a cDNA encoding a digestive
CCC	system antigen of the invention.
XXX	
SEQ	Sequence 421 BP; 148 A; 78 C; 127 G; 61 T; 7 other;
Query Match	22.4%; Score 34.2; DB 22; Length 421;
Best Local Similarity	64.6%; Pred. No. 0.71;
Matches	51; Conservative 0; Mismatches 28; Indels 0; Gaps
QY	52 AGAAGACACGAGCAAGCAGTTGCAGACCCTGGCAGATTACTTAATTAATTAATGAAT 111
Ddb	131 aaagggctcctcggggaagcagctacagaccgcgagcgtattgttgaaagctgctcaga 190
QY	112 AAAGACCTTGCGAAAAG 130
Ddb	191 aagggtctgggaagaag 209
RESULT 11	
ID	AARAA157603
XX	ID AAI57603 standard; cDNA; 421 BP.
AC	AAI57603;
XX	19-OCT-2001 (first entry)
XX	Human colorectal cancer antigen cDNA SEQ ID NO: 67.
DE	Human; colorectal cancer; colorectal cancer antigen; gene therapy; ss.
KW	Homo sapiens.
OS	WO200155350-A1.
PN	WO200155350-A1.
XX	02-AUG-2001.
XX	17-JAN-2001; 2001WO-US01350.
PPF	31-JAN-2000; 2000US-0179065.
XX	04-FEB-2000; 2000US-0180628.
RPR	24-FEB-2000; 2000US-0184664.
RPR	02-MAR-2000; 2000US-0186350.
RPR	16-MAR-2000; 2000US-0189874.
RPR	17-MAR-2000; 2000US-0190076.
RPR	18-APR-2000; 2000US-0198123.
RPR	19-MAY-2000; 2000US-0205515.
RPR	07-JUN-2000; 2000US-0209467.





```

RESULT_15
ABR82623
ID ABA82623 standard; DNA; 72049 BP.
XX AC
AC ABA82623;
DT DT
XT 25-JAN-2002 (first entry)
XX DE
DE Human HBM gene region b57d12_h_contig309G.
XX KW
KW Human; high bone mass; HTM gene; Zmaxl gene; chromosome 11; 11q13.3;
KK sequence tagged site; STS osteoporosis; osteopathic; gene therapy;
KW antisense therapy; vaccine; bone disorder; Paget's disease;
KW sclerostosis; osteomalacia; fibrous dysplasia; ds.
XX OS
OS Homo sapiens.
XX PN
PN WO200177327-A1.
PD PD
PD 18-OCT-2001.
XX XX
XX 21-JUN-2000; 200OWO-USI6951.
PR PR
PR 05-APR-2000; 2000US-0543771.
PR PR
PR 05-APR-2000; 2000US-0544398.
XX XX
XX (GENO-) GENOME THERAPEUTICS CORP.
PA PA
XX Carulli JP, Little RD, Recker RR, Johnson ML;
PI PI
PI WPI; 2001-657171/75.
DR DR
XX New high bone mass (HBM) and Zmaxl genes and proteins useful for
PT PT modulating bone mass for the treatment of e.g. osteoporosis -
PS PS Claim 51; Page 258-302; 443pp; English.
XX CC
CC The present invention describes the human Zmaxl gene and the high bone
CC mass (HBM) gene, which are found on chromosome 11q13.3. The Zmaxl and
CC HBM genes have osteopathic activities. The genes can be used in gene
CC therapy, antisense therapy and in the production of vaccines. They
CC can be used in the diagnosis and treatment of bone disorders including
CC osteoporosis, Paget's disease, sclerosis, osteomalacia and fibrous
CC dysplasia. ABR82038 to ABR82700 and AAG68168 to AAG68193 represent
CC sequences used in the exemplification of the present invention.
XX SQ
SQ Sequence 72049 BP; 15164 A; 19306 G; 20142 C; 17434 T; 3 other;

Query Match          21.3%; Score 32.6; DB 22; Length 72049;
Best Local Similarity 60.9%; Pred. No. 9.5;
Matches 53; Conservative 0; Mismatches 34; Indels 0; Gaps

QY      61   CAGCGAAGAAGCATTTCACGCCGGTGCAGATTAACCCTCATTAATAATTACTGAATAAGCACCTT 120
         ||||| | ||||| | ||||| | ||||| | ||||| | ||||| | ||||| 
Db    43349 ctg9gatcacagggtggtagccacgcgtctccttcgttttataactgctaagacctta 43408

QY     121   GCAGAAAGAAAGTAGTGCAAGAAGCTTTACT 147
        ||||| | ||||| | ||||| | ||||| | ||||| | ||||| | ||||| 
Db    43409 gataagaagaagtctgaagaagacttaact 43435


Search completed: August  2, 2002, 20:47:37
Job time: 25948 sec
```

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 2, 2002, 16:52:15 ; Search time 8940.62 seconds  
(without alignments)  
230.972 Million cell updates/sec

Title: US-08-973-363-5

Perfect score: 153

Sequence: 1 ATTTTACCTGATGACCCAGCA.....TGCAGAGCTTACTGTGGCA 153

Scoring table: IDENTIFY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

- 1: em\_estba:\*
- 2: em\_esthum:\*
- 3: em\_estin:\*
- 4: em\_estmu:\*
- 5: em\_estov:\*
- 6: em\_estpl:\*
- 7: em\_estro:\*
- 8: em\_hic:\*
- 9: gb\_est1:\*
- 10: gb\_est2:\*
- 11: gb\_hic:\*
- 12: gb\_gss:\*
- 13: em\_gss\_hum:\*
- 14: em\_gss\_inv:\*
- 15: em\_gss\_pln:\*
- 16: em\_gss\_vit:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	69	45.1	619	9	BB155356
2	69	45.1	660	9	BB461065
3	66	43.1	430	9	AL601246
C 4	66	43.1	547	9	AI890775
5	66	43.1	821	10	BF239967
6	66	43.1	866	9	AU125712
7	66	43.1	1028	10	BE895133
C 8	63.6	41.6	593	9	AL659353
9	62.6	40.9	645	9	AL644594
10	62.6	40.9	446	9	BH834922
C 11	59.4	38.8	686	9	AW907058
12	56.4	35.9	438	9	BH830730
C 13	50.4	32.9	327	9	AW996787
C 14	50.4	32.7	856	12	CNS04DVG
C 15	46.8	30.6	552	10	BI844947
C 16	46.8	30.6	574	10	BM185888
C 17	46.8	30.6	578	10	BM070535

C 18	46.8	30.6	608	10	BM024825
C 19	46.8	30.6	608	10	BM185005
C 20	44.4	29.0	642	9	AL658477
21	40.4	26.4	1013	12	CNS055XK
C 22	39.8	26.0	924	12	CNS04I26
C 23	39.8	26.0	1101	12	CNS04UQA
C 24	39.2	25.6	1156	12	CNS02JUV
C 25	39	25.5	812	12	CNS036YC
C 26	38.2	25.0	651	10	BM163891
C 27	38.2	25.0	724	12	CNS05DSJ
C 28	38.2	25.0	734	12	CNS056MU
C 29	38.2	25.0	923	12	CNS04J5I
30	38.2	25.0	963	12	CNS01X68
31	38.2	25.0	1015	12	CNS05KUD
C 32	38.2	25.0	1047	12	CNS04JPI
33	38.2	25.0	1050	12	CNS05N3J
C 34	38.2	25.0	1054	12	CNS05O6W
35	38.2	25.0	1074	12	CNS05IOP
36	38.2	25.0	1096	12	CNS05IC3
37	37	24.2	1122	12	CNS05T9J
C 38	36.8	24.1	904	12	CNS05990
39	36.6	23.9	967	12	CNS02BFS
C 40	36.6	23.9	1029	12	CNS05GSM
C 41	36.6	23.9	1036	12	CNS04VNY
C 42	36.6	23.9	1038	12	CNS05AVS
C 43	36.6	23.9	1051	12	CNS02H6W
C 44	36.6	23.9	1056	12	CNS05R92
C 45	36.6	23.9	1079	12	CNS05HIG

#### ALIGNMENTS

#### RESULT 1

BB155356 619 bp mRNA linear EST 18-OCT-2001  
musculus CDNA clone AL30024L16.3 similar to L10410 Mouse  
muscle-binding protein (CHD-1) mRNA, mRNA sequence.  
BB155356 2 GI:16268254  
EST.  
house mouse.  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 619)  
Arakawa.T., Carninci.P., Fukuda.S., Furuno.M., Hanagaki.T., Hara.A.,  
Hiramoto.K., Hori.F., Ishii.Y., Ito.M., Kawai.K., Kono.H., Kouda  
M., Koya.S., Matsuyama.T., Miyazaki.A., Nomura.K., Ohno.M.,  
Okazaki.Y., Okido.T., Saito.R., Sakai.C., Sakai.K., Sano.H., Sasaki  
D., Shibata.K., Shinagawa.A., Shiraki.T., Sogabe.Y., Suzuki.H.,  
Tagami.M., Tagawa.A., Takahashi.F., Takeda.Y., Tanaka.T., Toya.T.,  
Muramatsu.M. and Hayashizaki.Y.  
RIKEN Mouse ESTs (Arakawa.T., et al. 2001)  
Unpublished (2001)  
On Jun 29, 2000 this sequence version replaced gi:8811286.  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@sc.riken.go.jp/  
URL: http://genome.gsc.riken.go.jp/  
Carninci.P., Shibata.K., Hayashizaki.Y., Muramatsu.M. and Hayashizaki.Y.  
M., Kono.H., Okazaki.Y., Muramatsu.M. and Hayashizaki.Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. Genome Res. 10 (10), 1617-1630 (2000)  
wagi.K., Fujiwara.K., Inoue.K., Togawa.Y., Izawa.M., Ohara.E.,  
Watahiki.M., Yoneda.Y., Ishikawa.T., Ozawa.K., Tanaka.T., Matsuura

.S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
RIKEN integrated sequence analysis (RISA) system-384-format  
sequencing pipeline with 384 multicapillary sequencer. *Genome Res.*  
10 (11), 1757-1771 (2000)  
Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara  
, Y. and Hayashizaki, Y.  
Computer-based methods for the mouse full-length cDNA  
encyclopedia: real-time sequence clustering for construction of a  
nonredundant cDNA library. *Genome Res.* 11 (2), 281-289 (2001)  
Kondo, S., Shinagawa, A., Saito, T., Kiyosawa, H., Yamanaka, I., Aizawa  
, K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K. and  
Hayashizaki, Y.  
Computational Analysis of Full-Length Mouse cDNAs Compared with  
Human Genome Sequences *Mamm. Genome*. 12, 673-677 (2001)  
Please visit our web site (<http://genome.gsc.riken.go.jp/>) for  
further details.  
cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.

FEATURES	
SOURCE	

Location/Qualifiers:

```

1. 619
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone_lib="A130024L16"
/tissue_type="thymus"
/dev_stage="16 days neonate"
/lab_host="Dh108"
/notes="Site.1: Sali; Site.2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN, Division of Experimental Animal Research in RIKEN
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5',
GAGAGAGAGAGAGATCCACAGAGCTCTTTTTTTTTTTTTTNN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length
cap-trapper. cDNA went through one round of normalization
to Rot = 10.0 and subtraction to Rot = 195.0. Second
strand cDNA was prepared with the primer adapter of
sequence [5' GAGAGAGAGATCTCGAGTTAATTAATCCCCCCCCCC
3']. cDNA was cleaved with XhoI and BamHI. Vector: a
modified pBluescript KS(+) after bulk excision from Lambda
FIG. 1"

```

FLC I.	161	146
BASE COUNT	194 a	118 c

ORIGIN

Query Match 45.18; Score 69; DB 9; Length 619;

Best Local Similarity	80.2%	Pred. No. 2.1e-08;	
Matches	81;	Conservative	0; Mismatches 20; Indels 0; Gaps 0;

QY 52 AAGAAACCAAGCAAGCAGTTGCAGACCCGTGCAGATTACCTCATTAATTAAT 111

510 AAAAAACCACAGCAAAACAGTTTACAGACCCGTGCAGACTACCTCATCAAACTACTTAGC 569 Db

QY 112 AAAGACCTTGCAAGAAAGAAGTGCAAAGACTTACTGGTGC 152

570 AGAGATCTTGCAAAAAGAGAGGGCTCAGAGACTTTGTGGTGC 610

RESULT 2

BB461065

LOCUS

**DEFINITION** BB461065 RIKEN full-length enriched, 12 days embryo spinal ganglion  
Mus musculus cDNA clone D130070B13 3' similar to L10410 Mouse

VERSION BB461065.2 GI:16426612

KEYWORDS

**SOURCE**

ORGANISM Mus musculus

Mammalia; Eu

## REFERENCE

## AUTHORS

Hiramoto, K., Hori, F., Ishii, Y., Ito, M., Kawai, J., Konno, H., Kouda, M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, K., Ohno, M., Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Takeda, Y., Tanaka, T., Toyota, T., Muramatsu, M. and Hayashizaki, Y.

TITLE

JOURNAL

**COMMENT**

prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5', GAGAGAGATCTCGAGTTAATTAATATCCCGCCCCCCC 3']. cDNA was cleaved with BamHI and XhoI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."

BASE COUNT  
ORIGIN

Query Match 45.1%; Score 69; DB 9; Length 660;  
Best Local Similarity 80.2%; Pred. No. 2.1e-08;  
Matches 81; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY 52 AAGAACCCACAGGCAAGAGTTCGACACCGTGCAGATTACCTCAATAATTAAGTGAAT 111  
Db 502 AAAAAACCAAGCAAGAGTTCGACACCGTGCAGATTACCTCAATAATTAAGTGAAT 111  
Db 502 AAAAAACCAAGCAAGAGTTCGACACCGTGCAGATTACCTCAATAATTAAGTGAAT 111  
QY 112 AAGACCTTGCAGAAAGAGAGTTCGACACCGTGCAGATTACCTCAATAATTAAGTGAAT 152  
Db 562 AAGATCTTGCAGAAAGAGAGTTCGACACCGTGCAGATTACCTCAATAATTAAGTGAAT 152

RESULT 3  
AL601246  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

AL601246 430 bp mRNA linear EST 14-AUG-2001  
DKFZP313J1040\_r1 313 (synonym: hlcc2) Homo sapiens cDNA clone  
DKFZP313J1040 5', mRNA sequence.  
AL601246  
AL601246.1 GI:15164752  
EST  
human.

ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
AUTHORS Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.  
TITLE Bloeker, H., Boecker, M., Brandt, P., Mewes, W., Well, B. and Wiemann  
J.S. (1999)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Bloeker H  
MPS

Am Klopferspitz 18a D-62152 Martinsried, Germany  
This is the 5' sequence of the clone insert  
Cloned from S. Wiemann, Molecular Genome Analysis, German Cancer  
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;  
sequenced by GBF (National Research Centre for Biotechnology Ltd.,  
Braunschweig/Germany) within the cDNA sequencing consortium of the  
German Genome Project.  
No 5' sequence available.  
This clone (DKFZP313J1040) is available at the RZPD in Berlin.  
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059  
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

FEATURES  
source

1..430  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="DKFZP313J1040"  
/clone\_lib="313 (synonym: hlcc2)"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/note="Vector: pTriplex2; Site\_1: SfiIA; Site\_2: SfiIB;  
cDNA-collection"

BASE COUNT  
ORIGIN

Query Match 43.1%; Score 66; DB 9; Length 430;  
Best Local Similarity 87.8%; Pred. No. 1.3e-07;  
Matches 72; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 52 AAGAACCCACAGGCAAGAGTTCGACACCGTGCAGATTACCTCAATAATTAAGTGAAT 111  
Db 239 AAAAAACCAAGCAAGAGTTCGACACCGTGCAGATTACCTCAATAATTAAGTGAAT 298  
QY 112 AAGACCTTGCAGAAAGAGAGTTCGACACCGTGCAGATTACCTCAATAATTAAGTGAAT 133  
Db 239 AAGATCTTGCAGAAAGAGAGTTCGACACCGTGCAGATTACCTCAATAATTAAGTGAAT 320

RESULT 4  
AL890775/c  
LOCUS  
DEFINITION

AL890775 547 bp mRNA linear EST 07-MAR-2000  
wn95fll.x1 NCI\_CGAP\_Ut2 Homo sapiens cDNA clone IMAGE:2443725 3'  
Similar to SW:CHD1\_HUMAN O14646 CHROMODOMAIN-HELICASE-DNA-BINDING  
PROTEIN 1; mRNA sequence.  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE

AL890775  
AL890775.1 GI:5595939  
EST.  
human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
AUTHORS Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.  
TITLE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
Tumor Gene Index  
Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs@mail.nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: Life Technologies, Inc.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
www.bio.llnl.gov/bbrp/image/image.html  
Insert length: 1924 Std Error: 0.00  
Seq primer: -400P from Gibco  
High quality sequence stop: 418.  
Location/Qualifiers

1..547  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2443725"  
/clone\_lib="NCI\_CGAP\_Ut2"  
/tissue\_type="moderately-differentiated endometrial  
adenocarcinoma, 3 pooled tumors"  
/lab\_host="DH10B"  
/note="Organ: uterus; Vector: pCMV-SPORT6; Site\_1: SalI;  
Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 1.85 kb. Life Technologies catalog #:  
11539-012"

BASE COUNT 114 a 118 c 85 g 230 t  
ORIGIN

Query Match 43.1%; Score 66; DB 9; Length 547;  
Best Local Similarity 87.8%; Pred. No. 1.3e-07;  
Matches 72; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 52 AAGAACCCACAGGCAAGAGTTCGACACCGTGCAGATTACCTCAATAATTAAGTGAAT 111  
Db 357 AAAAAACCAAGCAAGAGTTCGACACCGTGCAGATTACCTCAATAATTAAGTGAAT 298  
QY 112 AAGACCTTGCAGAAAGAGAGTTCGACACCGTGCAGATTACCTCAATAATTAAGTGAAT 133  
Db 297 AAGATCTTGCAGAAAGAGAGTTCGACACCGTGCAGATTACCTCAATAATTAAGTGAAT 276

RESULT 5  
BF239967  
LOCUS

BF239967 821 bp mRNA linear EST 14-NOV-2000





```

source
1. .1028
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3921087"
/clone_lib="NIH_MGC_72"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: Sall; Cloned unidirectionally. Primer: Oligo dt. Average insert size 2 kb. Library constructed by Life Technologies."
BASE COUNT 387 a 205 c 238 g 198 t
ORIGIN
Query Match 43.1%; Score 66; DB 10; Length 1028;
Best Local Similarity 87.8%; Pred. No. 1.4e-07;
Matches 72; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 52 AAGAACCCAGGCAAGACAGTTCGACAGCCGTCGACAGTACCTCATTAATTAATGTAAT 111
DB 152 AAAAACCCAGGCAAGACAGTTCGACAGCCGTCGACAGTACCTCATTAATTAATGTAAT 111
QY 112 AAGACCTTGCAGAAAGAAAG 133
DB 212 AGAGATCTTGCAGAAAGAAAG 233

RESULT 8
AL659353/c
LOCUS AL659353 XGC-neurula Silurana tropicalis cDNA clone TNeu045e20 5', mRNA
DEFINITION AL659353 XGC-neurula Silurana tropicalis cDNA clone TNeu045e20 5', mRNA sequence.
ACCESSION AL659353
VERSION AL659353.1 GI:17672995
KEYWORDS EST.
SOURCE western clawed frog.
ORGANISM Silurana tropicalis
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipelidae; Xenopodinae; Silurana.
1 (bases 1 to 593)
Huckle,E., Taylor,R., Ashurst,J.L., Zorn,A.M. and Rogers,J.
Sanger Xenopus tropicalis EST project 2001 (10_2001)
Unpublished (2001)
Contact: Huckle E
Sanger Centre
Hinnton, Cambridgeshire, CB10 1SA, UK
Email: trop@sanger.ac.uk
TROPICALIS_SEQUENCE_ID: TNeu045e20.sp6
Sequencing primer: SP6
This sequence is from a Xenopus Gene Collection (XGC) library constructed by Aaron M. Zorn.
Location/Qualifiers
1. .593
/organism="Silurana tropicalis"
/db_xref="taxon:8364"
/clone="TNeu045e20"
/clone_lib="XGC-neurula"
/dev_stage="neurula"
/lab_host="Escherichia coli DH10B"
/notes="Vector: pCS107; Site_1: EcoRI; Site_2: NotI; cDNA was Oligo dt primed from 5ug of poly A+ RNA from neurula. EcoRI-NotI cut cDNA was then ligated into pCS107 with EcoRI at the 5' end and NotI at the 3' end."
BASE COUNT 119 a 134 c 114 g 226 t
ORIGIN
Query Match 41.6%; Score 63.6; DB 9; Length 593;
Best Local Similarity 76.5%; Pred. No. 5.7e-07;
Matches 78; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

source
1. .1028
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3921087"
/clone_lib="NIH_MGC_72"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: Sall; Cloned unidirectionally. Primer: Oligo dt. Average insert size 2 kb. Library constructed by Life Technologies."
BASE COUNT 387 a 205 c 238 g 198 t
ORIGIN
Query Match 43.1%; Score 66; DB 10; Length 1028;
Best Local Similarity 87.8%; Pred. No. 1.4e-07;
Matches 72; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 52 AAGAACCCAGGCAAGACAGTTCGACAGCCGTCGACAGTACCTCATTAATTAATGTAAT 111
DB 152 AAAAACCCAGGCAAGACAGTTCGACAGCCGTCGACAGTACCTCATTAATTAATGTAAT 111
QY 112 AAGACCTTGCAGAAAGAAAG 133
DB 212 AGAGATCTTGCAGAAAGAAAG 233

RESULT 8
AL659353/c
LOCUS AL659353 XGC-neurula Silurana tropicalis cDNA clone TNeu045e20 5', mRNA
DEFINITION AL659353 XGC-neurula Silurana tropicalis cDNA clone TNeu045e20 5', mRNA sequence.
ACCESSION AL659353
VERSION AL659353.1 GI:17672995
KEYWORDS EST.
SOURCE western clawed frog.
ORGANISM Silurana tropicalis
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipelidae; Xenopodinae; Silurana.
1 (bases 1 to 593)
Huckle,E., Taylor,R., Ashurst,J.L., Zorn,A.M. and Rogers,J.
Sanger Xenopus tropicalis EST project 2001 (10_2001)
Unpublished (2001)
Contact: Huckle E
Sanger Centre
Hinnton, Cambridgeshire, CB10 1SA, UK
Email: trop@sanger.ac.uk
TROPICALIS_SEQUENCE_ID: TNeu045e20.sp6
Sequencing primer: SP6
This sequence is from a Xenopus Gene Collection (XGC) library constructed by Aaron M. Zorn.
Location/Qualifiers
1. .593
/organism="Silurana tropicalis"
/db_xref="taxon:8364"
/clone="TNeu045e20"
/clone_lib="XGC-neurula"
/dev_stage="neurula"
/lab_host="Escherichia coli DH10B"
/notes="Vector: pCS107; Site_1: EcoRI; Site_2: NotI; cDNA was Oligo dt primed from 5ug of poly A+ RNA from neurula. EcoRI-NotI cut cDNA was then ligated into pCS107 with EcoRI at the 5' end and NotI at the 3' end."
BASE COUNT 119 a 134 c 114 g 226 t
ORIGIN
Query Match 41.6%; Score 63.6; DB 9; Length 593;
Best Local Similarity 76.5%; Pred. No. 5.7e-07;
Matches 78; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

```

```

QY 52 AAGAACCCAGGCAAGACAGTTCGACAGCCGTCGACAGTACCTCATTAATTAATGTAAT 111
DB 325 AAGAACCCAGGCAAGACAGTTCGACAGCCGTCGACAGTACCTCATTAATTAATGTAAT 266
QY 112 AAGACCTTGCAGAAAGAAAGTGCAGACCTTACTGTGTGCA 153
DB 265 AAGATCTGCTTAGGAAAGAGACCAAGACAGCTTCTAGTGA 224

RESULT 9
AL644594
LOCUS AL644594 XGC-egg Silurana tropicalis cDNA clone L1E1d12 5', mRNA
DEFINITION AL644594 XGC-egg Silurana tropicalis cDNA clone L1E1d12 5', mRNA sequence.
ACCESSION AL644594
VERSION AL644594.1 GI:16796719
KEYWORDS EST.
SOURCE western clawed frog.
ORGANISM Silurana tropicalis
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipelidae; Xenopodinae; Silurana.
1 (bases 1 to 645)
Huckle,E., Taylor,R., Ashurst,J.L., Zorn,A.M. and Rogers,J.
Sanger Xenopus tropicalis EST project 2001 (10_2001)
Unpublished (2001)
Contact: Huckle E
Sanger Centre
Hinnton, Cambridgeshire, CB10 1SA, UK
Email: trop@sanger.ac.uk
TROPICALIS_SEQUENCE_ID: L1E1d12.p1c
Sequencing primer: PIC
This sequence is from a Xenopus Gene Collection (XGC) library constructed by Aaron M. Zorn.
Location/Qualifiers
1. .645
/organism="Silurana tropicalis"
/db_xref="taxon:8364"
/clone="L1E1d12"
/clone_lib="XGC-egg"
/dev_stage="egg"
/lab_host="Escherichia coli XL1-blue"
/notes="Vector: pCS107; Site_1: EcoRI; Site_2: NotI; cDNA was Oligo dt primed from 5ug of poly A+ RNA from egg. EcoRI-NotI cut cDNA was then ligated into pCS107 with EcoRI at the 5' end and NotI at the 3' end"
BASE COUNT 222 a 125 c 156 g 141 t
ORIGIN
Query Match 41.6%; Score 63.6; DB 9; Length 645;
Best Local Similarity 76.5%; Pred. No. 5.8e-07;
Matches 78; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 52 AAGAACCCAGGCAAGACAGTTCGACAGCCGTCGACAGTACCTCATTAATTAATGTAAT 111
DB 477 AAGAACCCAGGCAAGACAGTTCGACAGCCGTCGACAGTACCTCATTAATTAATGTAAT 536
QY 112 AAGACCTTGCAGAAAGAAAGTGCAGACCTTACTGTGTGCA 153
DB 537 AAGATCTGCTTAGGAAAGAGACCAAGACAGCTTCTAGTGA 578

RESULT 10
BB834922
LOCUS BB834922 RIKEN full-length enriched, mammary gland RCB-0527
DEFINITION BB834922 RIKEN full-length enriched, mammary gland RCB-0527
JYg-MC(B) cDNA Mus musculus cDNA clone G93003J21 3', mRNA sequence.
ACCESSION BB834922
VERSION BB834922.1 GI:17013165

```



RESULT 13	AW996787	337 bp	mRNA	linear	EST 05-JUN-2000
LOCUS	QV3-BN0047-230200-102-d03	BN0047	Homo sapiens	CDNA	mRNA sequence.
DEFINITION	AW996787				
ACCESSION	AW996787.1	GI:8257021			
VERSION	EST.				
KEYWORDS	human.				
SOURCE	Homo sapiens				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
REFERENCE	1 (bases 1 to 337)				
AUTHORS	Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bata, G.S., Simpson, D.H., Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.				
TITLE	Shotgun sequencing of the human transcriptome with ORF expressed sequence tags				
JOURNAL	Proc. Natl. Acad. Sci. U.S.A.	97 (7),	3491-3496	(2000)	
MEDLINE	20202663				
COMMENT	Contact: Simpson A.J.G. Laboratory of Cancer Genetics Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil. Tel: +55-11-2704922 Fax: +55-11-2707001 Email: asimpson@ludwig.org.br This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (http://www.ludwig.org.br/scripts/gethtml2.pl?cli=st2-QV3-BN0047-2300-102-d03st3-2000-02-23st4-1) Seq primer: puc 18 forward High quality sequence start: 2 High quality sequence stop: 337. Location/Qualifiers 1..337 /organism="Homo sapiens" /db_xref="taxon:9606" /clone_lib="BN0047" /dev_stage="Adult" /note="Organ: breast_normal; Vector: puc18; Site: 1: SmaI; Site: 2: SmaI; A mini-library was made by cloning products derived from ORESPEC PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions." 83 a 59 c 72 g 123 t				
BASE COUNT	83 a	59 c	72 g	123 t	
ORIGIN					
Query Match	32.9%	Score 50.4;	DB 9;	Length 337;	
Best Local Similarity	84.1%	Pred. No. 0.0017;			
Matches	69;	Conservative 0;	Mismatches 11;	Indels 2;	Gaps 1;
QY	52	AGAAACACACAGCAAGCAGTTCGACACCGTCGACAGATTACCTCATTAATTAATCAAT	111		
Db	84	AAAAAACACACAGCAAGCAGTTCGACACCGTCGACAGATTACCTCATTAATTAATCAAT	111		
QY	112	AAGACCTTGTGCAAGAAAGAG	133		
Db	26	AGAGATCTTGTGCAAGAAAGAG	5		
RESULT 14	CNS04DVG	856 bp	DNA	linear	GSS 21-MAY-2000
LOCUS	CNS04DVG/c				
DEFINITION	Tetradodon nigroviridis genome survey sequence T7 end of clone 103k08 of library G from Tetradodon nigroviridis, genomic survey sequence.				

AL286261  
VERSION GI:8024707  
KEYWORDS GSS; genome survey sequence.  
SOURCE Tetraodon nigroviridis.  
ORGANISM Tetraodon nigroviridis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes;  
Tetraodontidae; Tetraodon.  
1 (bases 1 to 856)  
Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,  
Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and  
Weissenbach,J.  
Characterization and repeat analysis of the compact genome of the  
freshwater pufferfish Tetraodon nigroviridis

Unpublished  
2 (bases 1 to 856)  
Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,  
Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,  
Saurin,W. and Weissenbach,J.  
Human gene number estimate provided by genome wide analysis using  
Tetraodon nigroviridis DNA sequence

Unpublished  
3 (bases 1 to 856)  
Genoscope.  
Direct Submission  
Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases  
This sequence is a single read and was generated as part of a large  
scale clone-end sequencing project of the Tetraodon nigroviridis  
genome. For more information, please take a look at  
http://www.genoscope.cns.fr/Tetraodon.  
Location/Qualifiers  
1 . 856  
/organism="Tetraodon nigroviridis"  
/db\_xref="taxon:99883"  
/clone="103K08"  
/clone\_lib="G"  
/note="Genoscope sequence ID : COBG103BF04LP1-end : 77"

BASE COUNT 84 a 308 c 252 g 206 t 6 others  
ORIGIN

Query Match 32.7%; Score 50; DB 12; Length 856;  
Best Local Similarity 67.3%; Pred. NO. 0.0025;  
Matches 66; Conservative 3; Mismatches 29; Indels 0; Gaps 0;

QY 52 AAGAACACAGGCAAGCAGTTCAGACCCGTCAGATTACTCATTAATTAATCTGAAT 111  
||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
DB 656 AAGAGCCTCAGCCCAACAGTTCAGACCCGTCAGATTACTCATTAATTAATCTGAAT 111  
||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 112 AAGACCTTGCAAGAAAGAGTGCAGAACTTACTGG 149  
|| |||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
DB 596 AAGCACCTGGCCAGGAAGAGCCCAAGCAGGSGG 559  
||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 15  
BI844947/c  
LOCUS 532 bp mRNA linear EST 04-OCT-2001  
DEFINITION fq17e05.x1 zebrafish adult brain Danio rerio cDNA clone 4831593 3'  
similar to SW:CHD2\_HUMAN Q14647 CHROMODOMAIN-HELICASE-DNA-BINDING  
PROTEIN 2 ; , mRNA sequence.  
BI844947  
VERSION BI844947.1 GI:15957470  
KEYWORDS EST.  
SOURCE zebrafish.  
ORGANISM Danio rerio  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes  
; Cyprinidae; Danio.  
1 (bases 1 to 532)  
Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy  
S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood  
,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B.,

Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E.,  
Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.  
and Wilson,R.  
WashU zebrafish EST Project 1998  
Unpublished (1998)  
Other\_ESTs: fq17e05.y1  
Contact: Stephen L. Johnson  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: zbrafish@watson.wustl.edu  
CDNA Library Preparation: John Ngai. CDNA Library Arrayed by:  
Matthew Clark. DNA Sequencing by: Washington University Genome  
Sequencing Center Clone Distribution: Genome Systems, St. Louis,  
Missouri (web address: www.genomesystems.com) (email contact:  
info@genomesystems.com) and Research Genetics, Huntsville, Alabama  
(web address: www.resgen.com) (email contact: info@resgen.com) and  
ResourcenzentrumPrimatDatenbank, Berlin, Germany (web address:  
www.rzpd.de)  
Trace considered overall poor quality  
Seg primer: T7 from Gibco  
High quality sequence stop: 1.  
Location/Qualifiers  
1 . 532  
/organism="Danio rerio"  
/db\_xref="taxon:7955"  
/clone="4831593"  
/clone\_lib="zebrafish adult brain"  
/sex="mixed male and female"  
/tissue\_type="brain"  
/dev\_stage="adult"  
/lab\_host="E. coli DH10B"  
/note="vector: pZiPlox; Site:1: NotI; Site:2: SalI;  
Original library was constructed in lambdaZiPlox. Mass  
excision of the cDNA library was performed to yield  
pZiPlox plasmids. Insert check was done in original  
library."  
BASE COUNT 88 a 128 c 104 g 212 t  
ORIGIN

Query Match 30.6%; Score 46.8; DB 10; Length 532;  
Best Local Similarity 73.2%; Pred. NO. 0.016;  
Matches 60; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

QY 52 AAGAACACAGGCAAGCAGTTCAGACCCGTCAGATTACTCATTAATTAATCTGAAT 111  
||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
DB 349 AAGAACCTCAGGAAGAGCAGTTACAGATGCGAAGACAGACTACTTCTGAAGATGCTGAAG 290  
||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 112 AAGACCTTGCAAGAAAGAG 133  
||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
DB 289 AAGACCTGGAAGCCAGATG 268  
||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Search completed: August 2, 2002, 16:52:15  
Job time: 12216 sec

Mon Aug 5 11:51:55 2002

us-08-973-363-5.rst

Page 9



\_\_\_\_\_

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 2, 2002, 20:31:51 ; Search time 12674 Seconds  
(without alignments)  
10910.761 Million cell updates/sec

Title: US-08-973-363-10  
Perfect score: 6608  
Sequence: 1 CGGCTCGGCGACGAGCGC.....AACTAAACCAAAAAAACCC 6608

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

#### Database :

GenEmbl.\*

- 1: gb\_ba.\*
- 2: gb\_htg.\*
- 3: gb\_in.\*
- 4: gb\_lm.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_yi.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_hum.\*
- 18: em\_in.\*
- 19: em\_mu.\*
- 20: em\_om.\*
- 21: em\_or.\*
- 22: em\_ov.\*
- 23: em\_pat.\*
- 24: em\_ph.\*
- 25: em\_pl.\*
- 26: em\_ro.\*
- 27: em\_sts.\*
- 28: em\_un.\*
- 29: em\_vi.\*
- 30: em\_htg\_hum.\*
- 31: em\_htg\_inv.\*
- 32: em\_htg\_other.\*
- 33: em\_htg\_inv.\*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
------------	-------	-------------	--------	----	-------------

1	6608	100.0	6608	6	A58691	A58691 Sequence 10
2	6334	95.9	6872	5	AF004397	AF004397 Gallus ga
3	3847.8	58.2	5947	9	AF006513	AF006513 Homo sapi
4	3398.2	51.4	5349	10	MUSCHD1X	L10410 Mouse DNA-b
5	2538.2	38.4	2943	5	AF181826	AF181826 Gallus ga
6	2458	37.2	2754	5	AF181825	AF181825 Aegolius
7	2458	37.2	2754	5	AF181828	AF181828 Nymphicus
8	2422.8	36.7	2754	5	AF181824	AF181824 Aegolius
9	2406.8	36.4	2754	5	AF181827	AF181827 Nymphicus
10	2184.4	33.1	2292	5	DI4316	DI4316 delta-cryst
11	1522	23.0	7764	9	AF006514	AF006514 Homo sapi
12	1104.6	16.7	1316	6	A58696	A58696 Sequence 15
13	867.4	13.1	14530	2	AC019885	AC019885 Drosophila
14	867.4	13.1	153403	3	AC007765	AC007765 Drosophila
15	867.4	13.1	164920	3	AC092187	AC092187 Drosophila
16	867.4	13.1	306436	3	AE003581	AE003581 Drosophila
17	866.2	13.1	92442	3	DROCHDA	L72907 Drosophila
18	749.2	11.3	92783	3	AC084459	AC084459 Caenorhab
19	667.6	10.1	18029	3	CEH06001	Z92970 Caenorhab
20	604.6	9.1	101820	9	AC092372	AC092372 Homo sapi
21	604.6	9.1	143079	2	AC021449	AC021449 Homo sapi
22	604.6	9.1	219258	9	AC022121	AC022121 Homo sapi
23	603	9.1	134365	9	AC012624	AC012624 Homo sapi
24	592.2	9.0	145659	2	AC008531	AC008531 Homo sapi
25	552.4	8.4	635	5	AF059276	AF059276 Struthio
26	520.6	7.9	615	5	ECY12941	Y12941 Emberiza ci
27	520.6	7.9	615	5	FAY12942	Y12942 Ficedula al
28	519	7.9	615	5	FAY12939	Y12939 Ficedula al
29	512.6	7.8	615	5	ECY12940	Y12940 Emberiza ci
30	427.8	6.5	28491	8	SPAC3G6	U136499 S.pombe chr
31	395.6	6.0	19469	8	SPAC1783	U136499 S.pombe chr
32	393.8	6.0	67883	8	SCER229	AL18917 Saccharomyc
33	376.6	5.7	7981	8	SPHRE1	AX9021 S.pombe hrp
34	365.8	5.5	503	6	AX340692	AX340692 Sequence
35	325.2	4.9	4887	8	SCYOR304W	275212 S.cerevisia
36	323.2	4.9	5603	3	AF308444	AF308444 Caenorhab
37	321.6	4.9	5689	3	AF308445	AF308445 Caenorhab
38	318.6	4.8	3703	10	AF325920	AF325920 Mus muscu
39	317.8	4.8	3577	8	SCY8R245C	236114 S.cerevisia
40	314.4	4.8	2275	9	BC007347	BC007347 Homo sapi
41	312.8	4.7	6331	9	AF006515	AF006515 Homo sapi
42	312.8	4.7	6771	9	HSU81543	U91543 Homo sapien
43	311.2	4.7	6327	6	A56817	A56817 Sequence 1
44	311.2	4.7	6417	6	AX332238	AX332238 Sequence
45	311.2	4.7	6417	9	HSW12218	X86691 H.sapiens m

#### ALIGNMENTS

RESULT	1	A58691	Sequence 10 from Patent WO9639505.	DNA	linear	PAT 06-MAR-1998
LOCUS	A58691	6608 bp				
DEFINITION	A58691					
ACCESSION	A58691					
VERSION	A58691.1	GI:3714250				
KEYWORDS	unidentified.					
SOURCE	unclassified.					
ORGANISM	unclassified.					
REFERENCE	1 (bases 1 to 6608)					
AUTHORS	Griffiths, R. and Tiwari, B.					
TITLE	AVIAN GHD GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN BIRDS					
JOURNAL	Patent: WO 9639505-A 10 12-DEC-1996;					
COMMENT	ISIS INNOVATION (GB)					
FEATURES	Other publication AU 5906996 961224.					
SOURCE	Location/Qualifiers					
	1..6608					
	/organism="unidentified"					
	/db_xref="taxon:32644"					
BASE COUNT	2289 a 1207 c 1459 g 1653 t					
ORIGIN						

```
Query Match      100.0%; Score 6608; DB 6; Length 6608;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 6608; Conservative 0; Mismatches 0; Indels 0;
```

 $\ddot{O}$ 

Qy	1	CGG	GCTCGGG	CACGA	GCACG	CCGCCGCGGCACG	CAGGCTCGGGCCGGGGAGG	CCTGGC	60
Db	1	CGG	GCTCGGC	CACGA	GCACG	CCGCCGCGGCACG	CAGGCTCGGGCCGGGGAGG	CCTGGC	60
Qy	61	CCG	CCGACGCG	CACG	CACG	CTATTTGGG	CAAAAATCTTGGCC	ATCTGTAG	120
Db	61	CCG	CCGACGCG	CACG	CACG	CTATTTGGG	CAAAAATCTTGGCC	ATCTGTAG	120
Qy	121	CAAG	CTAAAGCG	CACTTACTT	CGAAAA	CATACGGAG	TACCAAGAGGGGATCTT	CTGACCTAC	180
Db	121	CAAG	CTAAAGCG	CACTTACTT	CGAAAA	CATACGGAG	TACCAAGAGGGGATCTT	CTGACCTAC	180
Qy	181	ACCT	TGTAACCT	GAGTGG	ACTTCTTTT	TAACTCTTTAA	CTCTTAATCTTCAATG	ATGGGCACA	240
Db	181	ACC	TGTAACCT	GAGTGG	ACTTCTTTT	TAACTCTTTAA	CTCTTAATCTTCAATG	ATGGGCACA	240
Qy	241	GTG	ATGAAGAAG	TGTAAG	AAAA	CACAGCAG	TGGAGAGTCAAG	CAGATGATCTG	300
Db	241	GTG	ATGAAGAAG	TGTAAG	AAAA	CACAGCAG	TGGAGAGTCAAG	CAGATGATCTG	300
Qy	301	GGT	CAGCTT	CAGGTTCT	TGGATCTGGT	TTCAGGCTCTG	GAAAGCAGTACCG	ATGGAAGTACGA	360
Db	301	GGT	CAGCTT	CAGGTTCT	TGGATCTGGT	TTCAGGCTCTG	GAAAGCAGTACCG	ATGGAAGTACGA	360
Qy	361	GCG	AGTCAGG	TAGCAGT	ACTCTG	AACTCTGTT	TCAGAGTCAGG	CACTCAATCCG	420
Db	361	GCC	AGTCAGG	TAGCAGT	ACTCTG	AACTCTGTT	TCAGAGTCAGG	CACTCAATCCG	420
Qy	421	AGT	CTGCACAT	CTAG	AGAGAA	GAACAAGTTCA	AGCTTAAACCTCCG	AAAGCTGACGGAT	480
Db	421	AGT	CTGCACAT	CTAG	AGAGAA	GAACAAGTTCA	AGCTTAAACCTCCG	AAAGCTGACGGAT	480
Qy	481	CTG	AGTTTTG	GGAAGTCC	AGGTC	CAAGCATACTT	CTGTACAGAGATCAG	CAGTGTCTAAGA	540
Db	481	CTG	AGTTTTG	GGAAGTCC	AGGTC	CAAGCATACTT	CTGTACAGAGATCAG	CAGTGTCTAAGA	540
Qy	541	AGC	AACAGCA	ACAGCA	AAAGCAG	CAATCAG	CACAGTGGTTTCA	GAAGGACNCTATCCA	600
Db	541	AGC	AACAGCA	ACAGCA	AAAGCAG	CAATCAG	CACAGTGGTTTCA	GAAGGACNCTATCCA	600
Qy	601	GTA	GTAAGATCT	CGCCG	ATGATTCG	TCAGTGA	AACTTAAGA	AAAAAGCATAAAGATG	660
Db	601	GTA	GTAAGATCT	CGCCG	ATGATTCG	TCCAGTGA	AACTTAAGA	AAAAAGCATAAAGATG	660
Qy	661	AAG	ACTGGCA	AAATGTC	AGGTC	AGGTC	AGTATCAGG	AACTGGTTCTGA	720
Db	661	AAG	ACTGGCA	AAATGTC	AGGTC	AGGTC	AGTATCAGG	AACTGGTTCTGA	720
Qy	721	CGG	AAAGTGGG	ATAAG	CAGTTGTG	AAAGAGTGA	ATCTGAC	TATCAGGCA	780
Db	721	CGG	AAAGTGGG	ATAAG	CAGTTGTG	AAAGAGTGA	ATCTGAC	TATCAGGCA	780
Qy	781	AAG	TCAAAAG	CCGTAA	ACCTCC	AAAGCA	GAATTAAG	CCAAAAAGTGG	840
Db	781	AAG	TCAAAAG	CCGTAA	ACCTCC	AAAGCA	GAATTAAG	CCAAAAAGTGG	840
Qy	841	GAC	AGAAAG	AGCA	ACTTGATTC	ATCAG	AGGAGGAGG	CAACATGAT	900
Db	841	GAC	AGAAAG	AGCA	ACTTGATTC	ATCAG	AGGAGGAGG	CAACATGAT	900
Qy	901	ATA	GAGAGG	ATCTCG	CCAGGCA	CAAGTGA	ATGTTTACA	AAAGCTCAAG	960
Db	901	ATA	GAGAGG	ATCTCG	CCAGGCA	CAAGTGA	ATGTTTACA	AAAGCTCAAG	960
Qy	961	CCA	AGACAG	ATTCAT	GATGATTTG	CTCG	GAAGTTTGG	GAGAGGATCTCC	1020
Db	961	CCA	AGACAG	ATTCAT	GATGATTTG	CTCG	GAAGTTTGG	GAGAGGATCTCC	1020





Db 4261 TAAAGAAGAAATAAAGAGTGAATTCCTCACCAACACCCCTCAGAAAAATCTGATGAAGATG 4320  
QY 4321 ATGAGGAGGAGATAACAAGGTAATGAATGAAATCTGAAAAATAAGAAAAATCTAAAA 4380  
Db 4321 ATGAGGAGGAGATAACAAGGTAATGAATGAAATCTGAAAAATAAGAAAAATCTAAAA 4380  
QY 4381 AAATTCATTTGCTGGATACCTCCAGTTCATATTAATCTGCAACCACTGGAACCAAGTTCCTATCT 4440  
Db 4381 AAATTCATTTGCTGGATACCTCCAGTTCATATTAATCTGCAACCACTGGAACCAAGTTCCTATCT 4440  
QY 4441 CAGAAGAATCTGAAGACCTCATCAGAGACATTTAGTGTGTCACAAAGAAATCAGGC 4500  
Db 4441 CAGAAGAATCTGAAGACCTCCATCAGAAGACATTTAGTGTGTCACAAAGAAATCAGGC 4500  
QY 4501 CTGTCAAGCAGACCTGAAACAGCTGGATAGACCAGAGAGGCCCTTCTGAAAGGAGC 4560  
Db 4501 CTGTCAAGCAGACCTGAACACCTGGATAGACCAGAGAGGCCCTTCTGAAAGGAGC 4560  
QY 4561 AGCTGGAACATACCTAGGCACTGTCTAATCAAAATTTGGGATCACAATACAGAATGCCTGA 4620  
Db 4561 AGCTGGAACATACCTAGGCACTGTCTAATCAAAATTTGGGATCACAATACAGAATGCCTGA 4620  
QY 4621 AGGAGTACACAAATCCCGAGCAATAAACAGTGGAGGAAATTTGTGGATTTTGTGT 4680  
Db 4621 AGGAGTACACAAATCCCGAGCAATAAACAGTGGAGGAAATTTGTGGATTTTGTGT 4680  
QY 4681 CCAAGTTTACAGAAATTTGATGCCAGAAAGCTGCACAAACTCTACAAACATGCCTGA 4740  
Db 4681 CCAAGTTTACAGAAATTTGATGCCAGAAAGCTGCACAAACTCTACAAACATGCCTGA 4740  
QY 4741 AGGCCAAGAGTCTAGCAACACAACTGACCAAAACATTTAGCAGCAATGTGAATACACATG 4800  
Db 4741 AGGCCAAGAGTCTAGCAACACAACTGACCAAAACATTTAGCAGCAATGTGAATACACATG 4800  
QY 4801 TAATCAGAATCCAGATGTGGAAGACTGAAGGAGACTACAAACCATGATGATAGTAGCA 4860  
Db 4801 TAATCAGAATCCAGATGTGGAAGACTGAAGGAGACTACAAACCATGATGATAGTAGCA 4860  
QY 4861 GGGACAGTTATTTCTGTAGACATTTATCACAATACCATGATCATCAAAAGACAGGC 4920  
Db 4861 GGGACAGTTATTTCTGTAGACATTTATCACAATACCATGATCATCAAAAGACAGGC 4920  
QY 4921 ATCAGGAGATGTTTACAGAAAAGTGACTCCAGGAAAAGGCCATTTACGCCCTTCAGTA 4980  
Db 4921 ATCAGGAGATGTTTACAGAAAAGTGACTCCAGGAAAAGGCCATTTACGCCCTTCAGTA 4980  
QY 4981 ATGGAAAAGATCAGAGACTGGGATCACTACAAAAGACAGACAGATACTACAGTGATA 5040  
Db 4981 ATGGAAAAGATCAGAGACTGGGATCACTACAAAAGACAGACAGATACTACAGTGATA 5040  
QY 5041 GTAACATAGAAAAGTTAGATGACCAAGGACAGACAGGTCAGGTCAGGTCGAAAGGAA 5100  
Db 5041 GTAACATAGAAAAGTTAGATGACCAAGGACAGACAGGTCAGGTCAGGTCGAAAGGAA 5100  
QY 5101 ACTTAAAGACACCCGGGTTCATTCAGATCACCGCTCCCATTCAGACACAGGATACACT 5160  
Db 5101 ACTTAAAGACACCCGGGTTCATTCAGATCACCGCTCCCATTCAGACACAGGATACACT 5160  
QY 5161 CAGATCACCGTTCACCTTCAGAAATACAGGCATCATAAATCTTCGAGAGATTAATAGATACC 5220  
Db 5161 CAGATCACCGTTCACCTTCAGAAATACAGGCATCATAAATCTTCGAGAGATTAATAGATACC 5220  
QY 5221 ACTCAGACTGGCAAAATGGACCACAGAGCTTCTGGTAGTGGCCCGAGGTCAACCACTAGATC 5280  
Db 5221 ACTCAGACTGGCAAAATGGACCACAGAGCTTCTGGTAGTGGCCCGAGGTCAACCACTAGATC 5280  
QY 5281 AGAGGTCTCCTTATGTTCAAGATCTCCCTTAGGACACAGATTCCTATTTGAACACTCAT 5340  
Db 5281 AGAGGTCTCCTTATGTTCAAGATCTCCCTTAGGACACAGATTCCTATTTGAACACTCAT 5340  
QY 5341 CAGATCACAAAAGTACACCTGAACATACATGGAGTAGCGGAGACATACAAAGACTGA 5400  
Db 5341 CAGATCACAAAAGTACACCTGAACATACATGGAGTAGCGGAGACATACAAAGACTGA 5400

QY 5401 CATTTTCTGGACCTTCTTTTAGCCATATACAGTAACTAAACACAGTAATTTGCCTTACAT 5460  
Db 5401 CATTTTCTGGACCTTCTTTTAGCCATATACAGTAACTAAACACAGTAATTTGCCTTACAT 5460  
QY 5461 GACTTGAAGAATAGTACCTGGATATTTCTATCAGTAGAGATTTGTTACTTCTTTCCAGGA 5520  
Db 5461 GACTTGAAGAATAGTACCTGGATATTTCTATCAGTAGAGATTTGTTACTTCTTTCCAGGA 5520  
QY 5521 TGCAGGTCCTATTATCCCAACAGAGAAAAATTTTGTATTAAAGTTTATGCTGCAC 5580  
Db 5521 TGCAGGTCCTATTATCCCAACAGAGAAAAATTTTGTATTAAAGTTTATGCTGCAC 5580  
QY 5581 TGTGCTGCAAAATGTTGTGGACCTTTTTTTTAAAGAAATGGAAGATGTTTACTTTTACAGG 5640  
Db 5581 TGTGCTGCAAAATGTTGTGGACCTTTTTTTTAAAGAAATGGAAGATGTTTACTTTTACAGG 5640  
QY 5641 GACCTCAACACTGCCCTTTTCAGACTGGATCTTACTATAAACTCTTCATGTCAAAGTGG 5700  
Db 5641 GACCTCAACACTGCCCTTTTCAGACTGGATCTTACTATAAACTCTTCATGTCAAAGTGG 5700  
QY 5701 TTCTAGGCTGAACACAGATTAATTAATTTGTTGAAATGAACACTTAAACACTGACCTGTG 5760  
Db 5701 TTCTAGGCTGAACACAGATTAATTAATTTGTTGAAATGAACACTTAAACACTGACCTGTG 5760  
QY 5761 CTTATGTTTTCAGGAAAAGATGGGGATTTATTTTGTGTTTATTTCTTGGTAGAGAACTCTC 5820  
Db 5761 CTTATGTTTTCAGGAAAAGATGGGGATTTATTTTGTGTTTATTTCTTGGTAGAGAACTCTC 5820  
QY 5821 AAGGACTTTGTTCACTTTCCAAAGCTACTGTTTACATTTGTACACTGCGACCACTTGCC 5880  
Db 5821 AAGGACTTTGTTCACTTTCCAAAGCTACTGTTTACATTTGTACACTGCGACCACTTGCC 5880  
QY 5881 GCTTTTCATCACAAAGCTTGAATATTTAAATCTCTGTACCTACAGTTCTGTAATAGCCAGGA 5940  
Db 5881 GCTTTTCATCACAAAGCTTGAATATTTAAATCTCTGTACCTACAGTTCTGTAATAGCCAGGA 5940  
QY 5941 TTTCTCCTGTTGTGATCAGTTATAATGCTTTTTTATGAAACAAACAAACAAACAAACAA 6000  
Db 5941 TTTCTCCTGTTGTGATCAGTTATAATGCTTTTTTATGAAACAAACAAACAAACAAACAA 6000  
QY 6001 CAATTAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAA 6060  
Db 6001 CAATTAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAA 6060  
QY 6061 AAATGAGCTTTTTCGCTCAGGCTTTTTTGGCTGTTTCTTTCCCAACAACTCAGGCTC 6120  
Db 6061 AAATGAGCTTTTTCGCTCAGGCTTTTTTGGCTGTTTCTTTCCCAACAACTCAGGCTC 6120  
QY 6121 TCTTTTCACAAAGTCAGTATACATGTTTAAATTAATAATATCTCGATGGAATCAGAT 6180  
Db 6121 TCTTTTCACAAAGTCAGTATACATGTTTAAATTAATAATATCTCGATGGAATCAGAT 6180  
QY 6181 GTAAAAATGGGGAAGGAATATTTTATCCATTTAGTGTCTCTTTTATTTGATGACTTT 6240  
Db 6181 GTAAAAATGGGGAAGGAATATTTTATCCATTTAGTGTCTCTTTTATTTGATGACTTT 6240  
QY 6241 TACATACCTGTTTTCGTTGTTTATTTTATTTTCTATTAACTGTGAGTGTG 6300  
Db 6241 TACATACCTGTTTTCGTTGTTTATTTTATTTTCTATTAACTGTGAGTGTG 6300  
QY 6301 TGATGTTGTAATGAACAGTGAGAAATCCCACTCTAAACTGTGCCCTGGAAGCTTTTC 6360  
Db 6301 TGATGTTGTAATGAACAGTGAGAAATCCCACTCTAAACTGTGCCCTGGAAGCTTTTC 6360  
QY 6361 AGGTGATTTGGTTTAAAGAGGAAGTGTCTATAGGTGAACACTTCAAAACCCAGATCA 6420  
Db 6361 AGGTGATTTGGTTTAAAGAGGAAGTGTCTATAGGTGAACACTTCAAAACCCAGATCA 6420  
QY 6421 GCCAAGTTCATTTGTAATCCATTTGTTTCCCTCTTTTAACTGGGCAATAATGTCAAAT 6480  
Db 6421 GCCAAGTTCATTTGTAATCCATTTGTTTCCCTCTTTTAACTGGGCAATAATGTCAAAT 6480



QY 781 AAGTCAAAAGCCGTAAACCTTCCAAGCAGAAATTAAGCCAAAAGTGGGAAAAGAGCACAG 840  
DB 781 AAGTCAAAAGCCGTAAACCTTCCAAGCAGAAATTAAGCCAAAAGTGGGAAAAGAGCACAG 840  
QY 841 GACAGAGAAGAGGACACTTGATTCATCAGAGAGAGGAGGACGATGATGAAGATTATG 900  
DB 841 GACAGAGAAGAGGACACTTGATTCATCAGAGAGAGGAGGAGGACGATGATGAAGATTATG 900  
QY 901 ATAAGAGAGGATCTCGTCGCCAGGCAACAGTGAATGTTACTTACAAAGAGCTGAAAGAAA 960  
DB 901 ATAAGAGAGGATCTCGTCGCCAGGCAACAGTGAATGTTACTTACAAAGAGCTGAAAGAAA 960  
QY 961 CCAAGACAGATCTGATGATTTGCTGGAAGTTGTTGAGAGGATGTCGCCACAGACTGAAG 1020  
DB 961 CCAAGACAGATCTGATGATTTGCTGGAAGTTGTTGAGAGGATGTCGCCACAGACTGAAG 1020  
QY 1021 AAGATGAATTTGAACTATAGAGAAGTTTATGGACAGTGAATTTGGCCGAAAAGGAGCCA 1080  
DB 1021 AAGATGAATTTGAACTATAGAGAAGTTTATGGACAGTGAATTTGGCCGAAAAGGAGCCA 1080  
QY 1081 CTGGTGCTCAACACCACTATGCGGTTGAGGCAGATGGTGACCAAAATGCTGGGTTTG 1140  
DB 1081 CTGGTGCTCAACCACTATGCGGTTGAGGCAGATGGTGACCAAAATGCTGGGTTTG 1140  
QY 1141 AAAAGTCAAAAGGAGCTGGGAGAAATACAGTATCTTATTAATGGAAGGCTGGTCAACA 1200  
DB 1141 AAAAGTCAAAAGGAGCTGGGAGAAATACAGTATCTTATTAATGGAAGGCTGGTCAACA 1200  
QY 1201 TCCATACACTTGGGAACHTGAAGAACGCTGAAGCAACAAATGTTAAAGGAATGAACA 1260  
DB 1201 TCCATACACTTGGGAACHTGAAGAACGCTGAAGCAACAAATGTTAAAGGAATGAACA 1260  
QY 1261 AACTGGCAACTACAAGAAAAGGATCAGGAGACAAAACGCTGGCTGAAAATGCTTCTC 1320  
DB 1261 AACTGGCAACTACAAGAAAAGGATCAGGAGACAAAACGCTGGCTGAAAATGCTTCTC 1320  
QY 1321 CAGAAGATGGAAATATATACTGCCAGGAGGCTTACAGATGATCTGCACAAACAAT 1380  
DB 1321 CAGAAGATGGAAATATATACTGCCAGGAGGCTTACAGATGATCTGCACAAACAAT 1380  
QY 1381 ATCAAAATAGTGGAAAGAAATTAATGCTCATTAATCAAAAGTCAAGAGTGGTTATCCGG 1440  
DB 1381 ATCAAAATAGTGGAAAGAAATTAATGCTCATTAATCAAAAGTCAAGAGTGGTTATCCGG 1440  
QY 1441 ACTACTATTGCAAAATGSCAGGGTCTGCTTACTCAGAACTGAGATGATCTGCACAAACAAT 1500  
DB 1441 ACTACTATTGCAAAATGSCAGGGTCTGCTTACTCAGAACTGAGATGATCTGCACAAACAAT 1500  
QY 1501 TCATTGCCAAAAGTTTCAGGCACGATTCATGATCAGTATTTTACAGAAATCAATCCAAAG 1560  
DB 1501 TCATTGCCAAAAGTTTCAGGCACGATTCATGATCAGTATTTTACAGAAATCAATCCAAAG 1560  
QY 1561 CTACTCCCTTTAAGGACTGCAAGGTTCTAAACAGAGACCAAGATTTGTTGCACTGAAGA 1620  
DB 1561 CTACTCCCTTTAAGGACTGCAAGGTTCTAAACAGAGACCAAGATTTGTTGCACTGAAGA 1620  
QY 1621 AGCAACCATCTTACATTTGGAGGACATGAAAGTCTGGAGTTAAGAGATTATCAGTTAAATG 1680  
DB 1621 AGCAACCATCTTACATTTGGAGGACATGAAAGTCTGGAGTTAAGAGATTATCAGTTAAATG 1680  
QY 1681 GATTGAATTTGGCTCGCTCATTCATGTTGCAAGGAATAGTTGTTGTCAGATGAAGA 1740  
DB 1681 GATTGAATTTGGCTCGCTCATTCATGTTGCAAGGAATAGTTGTTGTCAGATGAAGA 1740  
QY 1741 TGGGCTCGGGTAAAACAAATACAAACAATTTCTTTTCTGAACTACCTGTTTCATGAACATC 1800  
DB 1741 TGGGCTCGGGTAAAACAAATACAAACAATTTCTTTTCTGAACTACCTGTTTCATGAACATC 1800  
QY 1801 AACTGTATGSCCCTTTCTTCGCGGTGCCACTTTCTACCTTTGACATCTTGGCAAGAG 1860  
DB 1801 AACTGTATGSCCCTTTCTTCGCGGTGCCACTTTCTACCTTTGACATCTTGGCAAGAG 1860

QY 1861 AGATTCAAACTTGGGCTCCTCAGATGAATGCTGTAGTTACTTTAGGAGATATTAACTAGTA 1920  
DB 1861 AGATTCAAACTTGGGCTCCTCAGATGAATGCTGTAGTTACTTTAGGAGATATTAACTAGTA 1920  
QY 1921 GAAATATGATAAGGACTCATGAATGGATGCATCCACAGACTAAACGATTAAAGCTTAAACA 1980  
DB 1921 GAAATATGATAAGGACTCATGAATGGATGCATCCACAGACTAAACGATTAAAGCTTAAACA 1980  
QY 1981 TACTTCTGACGACATATGAAATTTTACTGAAGGATAAGTCATTCCTTGGTGGTCTCAATT 2040  
DB 1981 TACTTCTGACGACATATGAAATTTTACTGAAGGATAAGTCATTCCTTGGTGGTCTCAATT 2040  
QY 2041 GGGCATTTCAATAGAGTTGATGAAGCTCATCGTTTAAAAAATGATGACTCTCTCTGTACA 2100  
DB 2041 GGGCATTTCAATAGAGTTGATGAAGCTCATCGTTTAAAAAATGATGACTCTCTCTGTACA 2100  
QY 2101 GGACTTTAATAGACTTTAAGTCCAACCATCGACTTCTGATTACTGGNACCCACTGCAAA 2160  
DB 2101 GGACTTTAATAGACTTTAAGTCCAACCATCGACTTCTGATTACTGGNACCCACTGCAAA 2160  
QY 2161 ATTCCCTCAAAAGAGCTGTGGTCTTGTTCATTTTCATCATGCCAGAAAAATTTTCTCTCT 2220  
DB 2161 ATTCCCTCAAAAGAGCTGTGGTCTTGTTCATTTTCATCATGCCAGAAAAATTTTCTCTCT 2220  
QY 2221 GGGAAATTTTGAAGAGGAGCATGGCAAAAGGAGATATGTTATGCAAGTCTTTCACA 2280  
DB 2221 GGGAAATTTTGAAGAGGAGCATGGCAAAAGGAGATATGTTATGCAAGTCTTTCACA 2280  
QY 2281 AGAGCTTGAACCATTTTACTAAGAGAGTTTAAAAAGATGTAGAAAAGTCTTTACCTG 2340  
DB 2281 AGAGCTTGAACCATTTTACTAAGAGAGTTTAAAAAGATGTAGAAAAGTCTTTACCTG 2340  
QY 2341 CTAAAGCTTGAACAAATTTCTGAGGATGGAATGATGCAATTCGACAGAAACAATATTACAAGT 2400  
DB 2341 CTAAAGCTTGAACAAATTTCTGAGGATGGAATGATGCAATTCGACAGAAACAATATTACAAGT 2400  
QY 2401 GGATTTTACAAGGAATTTATAAGCCCTCAGTAAAGGTTCAAAAAGGAGTACCTCAGGCT 2460  
DB 2401 GGATTTTACAAGGAATTTATAAGCCCTCAGTAAAGGTTCAAAAAGGAGTACCTCAGGCT 2460  
QY 2461 TTTCTGAACATTTATGATGGAACTTTAAGAAGTGTGTAAACCATTCCTACCTCATTAAGCCAC 2520  
DB 2461 TTTCTGAACATTTATGATGGAACTTTAAGAAGTGTGTAAACCATTCCTACCTCATTAAGCCAC 2520  
QY 2521 CAGATGATAATGAATTTCTATAATTAACAGSAGGCCCTTACAGCATTTTGATACGTAGCAGCG 2580  
DB 2521 CAGATGATAATGAATTTCTATAATTAACAGSAGGCCCTTACAGCATTTTGATACGTAGCAGCG 2580  
QY 2581 GGAACTAATCTCTTCTTGACAAGCTACTGATTCGTCTGGAGAACGTTGGCAACAGAGTTT 2640  
DB 2581 GGAACTAATCTCTTCTTGACAAGCTACTGATTCGTCTGGAGAACGTTGGCAACAGAGTTT 2640  
QY 2641 TGATTTTCTCTCAGATGGTGAGGATCTGCACATCCTAGCAGAAATATCTGAAGTATCGCC 2700  
DB 2641 TGATTTTCTCTCAGATGGTGAGGATCTGCACATCCTAGCAGAAATATCTGAAGTATCGCC 2700  
QY 2701 AGTTTCCCTTCCAGAGACTTGATGGATCAATAAAAGGGAAATTTAGSAAAGCAAGCACTGG 2760  
DB 2701 AGTTTCCCTTCCAGAGACTTGATGGATCAATAAAAGGGAAATTTAGSAAAGCAAGCACTGG 2760  
QY 2761 ATCATTTTCAATGCAGAAAGATCAGAGATTTCTGTTTTTTTACTGTCTACAGAGCTGGAG 2820  
DB 2761 ATCATTTTCAATGCAGAAAGATCAGAGATTTCTGTTTTTTTACTGTCTACAGAGCTGGAG 2820  
QY 2821 GATTAGTATTAACTTGGCATCTGCTGACACTGTAGTATTTTGTATTTCTGACTGGAATC 2880  
DB 2821 GATTAGTATTAACTTGGCATCTGCTGACACTGTAGTATTTTGTATTTCTGACTGGAATC 2880  
QY 2881 CACAGAATGATCTGCAGGCACAGCGGAGAGCTCATAGAATTTGGACAGAAACAGAGTTA 2940  
DB 2881 CACAGAATGATCTGCAGGCACAGCGGAGAGCTCATAGAATTTGGACAGAAACAGAGTTA 2940  
QY 2941 ATATTTTATCGGCTAGTCAAAAAGGATCAGTAGAAGATATTTCTTGAAGAGCCNAGA 3000

|||||ATATTATCGGCTAGTCAAAAGGATCAGTAAAGAGATATTTCTGAAAGAGCCAAAG 3000  
Db 2941  
QY 3001 AGAAGATGGTGTAGACCATTTAGTAATTCAGAGAAATGGACACGACGAGAAAACTGTTG 3060  
Db 3001 AGAAGATGGTGTAGACCATTTAGTAATTCAGAGAAATGGACACGACGAGAAAACTGTTG 3060  
QY 3061 TGCATACAGGTTCAACTCCATCAAGCTCTACACCTTTTAAATAAAGAGAGTTATCAGCTA 3120  
Db 3061 TGCATACAGGTTCAACTCCATCAAGCTCTACACCTTTTAAATAAAGAGAGTTATCAGCTA 3120  
QY 3121 TTTTGAAGTTTGGTGTCTGAGGAACCTTTTAAAGAACCTTGAAGGAGAGAAACAGAGGCC 3180  
Db 3121 TTTTGAAGTTTGGTGTCTGAGGAACCTTTTAAAGAACCTTGAAGGAGAGAAACAGAGGCC 3180  
QY 3181 AGGAATGGATAGATGAATCTTGAAGAGAGCTGAACCTGGGAAATGAGCCAGGTC 3240  
Db 3181 AGGAATGGATAGATGAATCTTGAAGAGAGCTGAACCTGGGAAATGAGCCAGGTC 3240  
QY 3241 CATTTGACTGTAGGGATGAGTTGCTTTACAGTTCAAGGTGGCGAACTTTTCCCAATG 3300  
Db 3241 CATTTGACTGTAGGGATGAGTTGCTTTACAGTTCAAGGTGGCGAACTTTTCCCAATG 3300  
QY 3301 ATGAAGATGATTTGAGTTGGAACCAAGAAATTCAGAAATTTGGGAAGAAATCATCC 3360  
Db 3301 ATGAAGATGATTTGAGTTGGAACCAAGAAATTCAGAAATTTGGGAAGAAATCATCC 3360  
QY 3361 CAGAAATCCCAACGAGAGATAGAGGAGGAGAAAGACAAAGAACTTGAAGAAATAT 3420  
Db 3361 CAGAAATCCCAACGAGAGATAGAGGAGGAGAAAGACAAAGAACTTGAAGAAATAT 3420  
QY 3421 ACATGCTCCCGAGGATGAGAACTGTCAAAACAGATCAGCTTTATGGAGTCAAGGAA 3480  
Db 3421 ACATGCTCCCGAGGATGAGAACTGTCAAAACAGATCAGCTTTATGGAGTCAAGGAA 3480  
QY 3481 GACGCTAGGAGCAGAGATATCTGGATCTGATGATCTCCAGAAATATTAAGGATTTAGTG 3540  
Db 3481 GACGCTAGGAGCAGAGATATCTGGATCTGATGATCTCCAGAAATATTAAGGATTTAGTG 3540  
QY 3541 GGCACAAAAGCGTGAAGACCTGGAACCTTCTCGAGAAATATTAAGGATTTAGTG 3600  
Db 3541 GGCACAAAAGCGTGAAGACCTGGAACCTTCTCGAGAAATATTAAGGATTTAGTG 3600  
QY 3601 ATGAGAGATCAGCGGTTTATCAGAGTTACAGAAATTTGGTGGCCCTCTGGAAGGT 3660  
Db 3601 ATGAGAGATCAGCGGTTTATCAGAGTTACAGAAATTTGGTGGCCCTCTGGAAGGT 3660  
QY 3661 TAGATGCTGTAGATAGATGCTGAACCTGGTTGATAATCTGAGACAGACCTTTAGACGTT 3720  
Db 3661 TAGATGCTGTAGATAGATGCTGAACCTGGTTGATAATCTGAGACAGACCTTTAGACGTT 3720  
QY 3721 TGGGTGAACCTGTACATATGATGATTAAGGCTTTTAAAGGACAAATTCATCTGGACAAG 3780  
Db 3721 TGGGTGAACCTGTACATATGATGATTAAGGCTTTTAAAGGACAAATTCATCTGGACAAG 3780  
QY 3781 AAAGACAGAGGTAGACTTGGGAAAGTTTAAAGGCCCAAGCTTTGGAATCTCAGAGTGC 3840  
Db 3781 AAAGACAGAGGTAGACTTGGGAAAGTTTAAAGGCCCAAGCTTTGGAATCTCAGAGTGC 3840  
QY 3841 AGGTGAATGCAAACTAGTCTATCTCAGAGAGAGAGTGGCACCCTGCACAAATCCA 3900  
Db 3841 AGGTGAATGCAAACTAGTCTATCTCAGAGAGAGAGTGGCACCCTGCACAAATCCA 3900  
QY 3901 TTCCTTCAGATCAGAAAGAAAGATATGTCATCCATGCCACAGGAGCTGCTC 3960  
Db 3901 TTCCTTCAGATCAGAAAGAAAGATATGTCATCCATGCCACAGGAGCTGCTC 3960  
QY 3961 ACTTCGATATGATTTGGGTAAGAGATGATTCGAATCTGTTAGTGGGATCTATGAAT 4020  
Db 3961 ACTTCGATATGATTTGGGTAAGAGATGATTCGAATCTGTTAGTGGGATCTATGAAT 4020  
QY 4021 ATGGCTATGCGAGCTGGGAAATGATAAAATGGATGCCAGATCTCAGCTTTAACACAGAGA 4080  
|||||

Db 4021 ATGGCTATGCGAGCTGGGAAATGATAAAATGGATCCAGATCTCAGCTTTAACACAGAGA 4080  
QY 4081 TTTTACCTGTATGTCAGACAAGAACCCCGAGCAAAAGCAGCTACAGACCCGTCGAGACT 4140  
Db 4081 TTTTACCTGTATGTCAGACAAGAACCCCGAGCAAAAGCAGCTACAGACCCGTCGAGACT 4140  
QY 4141 ACCTCATTAATTTACTGTAATAAGACCTTGCAGAAAGAAAGACACAAAGGCTTTGCTG 4200  
Db 4141 ACCTCATTAATTTACTGTAATAAGACCTTGCAGAAAGAAAGACACAAAGGCTTTGCTG 4200  
QY 4201 CAGGCAATTCAGAGAGAGAGACAAAGAAATAAGAAATAAGATGAAGGCTTCAAAA 4260  
Db 4201 CAGGCAATTCAGAGAGAGAGACAAAGAAATAAGAAATAAGATGAAGGCTTCAAAA 4260  
QY 4261 TAAAGAAGAAATAAGAGTGTATTTTCACACACACCCCTCAGAAATCTGATGAAGTG 4320  
Db 4261 TAAAGAAGAAATAAGAGTGTATTTTCACACACACCCCTCAGAAATCTGATGAAGTG 4320  
QY 4321 ATGAGGAGGAGGATAAC----- 4337  
Db 4321 ATGAGGAGGAGGATAACAGGATGAGATGTTTTCAGTGAACATCTACATAAAAAATAA 4380  
QY 4338 ----- 4337  
Db 4381 AAACAGAAAAAGAAATGAAGAAAGCCTGAGCCAGATATTTGGTATAAAGAAAGAGCTG 4440  
QY 4338 ----- 4337  
Db 4441 AAGAAAAAGAGAGACAAAGAGAGAAATAAAAGGGAAATTCGAAAGGGAGAAAAAG 4500  
QY 4338 ----- 4337  
Db 4501 AAAAGAGGATTAAGAAAGATTAAGACAAAAAGATAATAAGAAAGAGAGAAAAAAG 4560  
QY 4338 -----AAGTTAAATGAATGAAT 4356  
Db 4561 TAAAGAAATCCACACAGAAAGAAAGAAAGAAAGTGAAGAGAGAGAGTAAATGAATGAAT 4620  
QY 4357 CTGAAATAAAGAAAAATCTAAAAATTTCCATTTGCTGGATCTACCCAGTTTCAATTTACTG 4416  
Db 4621 CTGAAATAAAGAAAAATCTAAAAATTTCCATTTGCTGGATCTACCCAGTTTCAATTTACTG 4680  
QY 4417 CAACAGTGAACAGTTCTCTCTCAGAGAAATCTGAAGAACTCCATCAGAGAGACATTTA 4476  
Db 4681 CAACAGTGAACAGTTCTCTCTCAGAGAAATCTGAAGAACTCCATCAGAGAGACATTTA 4740  
QY 4477 GTGTGTCAAGAAAGAAATGAGGCTGTCAAGAGCAGACTGAAACAGCTGGATAGACCAG 4536  
Db 4741 GTGTGTCAAGAAAGAAATGAGGCTGTCAAGAGCAGACTGAAACAGCTGGATAGACCAG 4800  
QY 4537 AGAAGGGCTTTCTGAAAGGGAGGAGCTGGAACATCTAGGAGCTGTCTTAATCAAAATTTG 4596  
Db 4801 AGAAGGGCTTTCTGAAAGGGAGGAGCTGGAACATCTAGGAGCTGTCTTAATCAAAATTTG 4860  
QY 4597 GGGATCACATTACAGAAATGCTGAAAGGAGTACACAAATCCCGAGCAAAATAAAACAGTGA 4656  
Db 4861 GGGATCACATTACAGAAATGCTGAAAGGAGTACACAAATCCCGAGCAAAATAAAACAGTGA 4920  
QY 4657 GGAATAATTTGCTGGATTTTGTGTTCCAAAGTTTACAGAAATTTGATGCCAGAAAGCTGCACA 4716  
Db 4921 GGAATAATTTGCTGGATTTTGTGTTCCAAAGTTTACAGAAATTTGATGCCAGAAAGCTGCACA 4980  
QY 4717 AACTCTCAAAACATGCAATCAAAAAGCGCCAGAGCTCTCAGCAACATGACCAAAACA 4776  
Db 4981 AACTCTCAAAACATGCAATCAAAAAGCGCCAGAGCTCTCAGCAACATGACCAAAACA 5040  
QY 4777 TTAGACGCAATTTGAATACATGTAAATCAAGAAATCCAGATGTGGAAAGACTGAAGGAGA 4836  
Db 5041 TTAGACGCAATTTGAATACATGTAAATCAAGAAATCCAGATGTGGAAAGACTGAAGGAGA 5100  
QY 4837 CTACAAACCATGATGATAGTAGAGGAGGACAGTTATTTCTCTGTATGACATTTTATCACAAT 4896  
Db 5101 CTACAAACCATGATGATAGTAGAGGAGGACAGTTATTTCTCTGTATGACATTTTATCACAAT 5160







QY 1679 TGGATTGAATGGCTCGCTCAATTCATGGTGCAAGGAAATAGTGTATCTTCGACATGA 1738  
Db 1624 TGGTTTAAATGGCTGCTCATTTCTGGTGCAAGGAAATAGTGTGACATCTCGCTGATGA 1683  
QY 1739 AATGGGTCTGGGTAAACAATACAAACAATTTCTTTCTGCAACTACTGTTTTCATGAACA 1798  
Db 1684 AATGGGCTTTGGAAAAACAATACAGACGATCTCATTTCTGCAATATTTGTTTTCATGAACA 1743  
QY 1799 TCAACTGTATGGCCCTTTCTTCTCGCGCGCCACATTTCTACCTTGACATCTTTGGCAAG 1858  
Db 1744 TCAATATATGGACCTTTTTTATGGTAGTACCCTCTCCACCTTACTTCTCGGCAAG 1803  
QY 1859 AGAGATTCAAACTTGGGCTCCTCAGATGAATGCTGTAGTTTACTTATGAGGATATAACTAG 1918  
Db 1804 GGAATTCAGACTTGGCTTCTCAATGAATGCTGTGGTTTATTTAGGTGACATTAACAG 1863  
QY 1919 TAGAATATGATTAAGGACTCATGAATGGATGCAATCCACAGACATTAACGATTAAGTTTAA 1978  
Db 1864 CAGAAACATGATTAAGAACTCATGAATGGAGCGCATCATCAGACCAACCGGTTAAAAATTTAA 1923  
QY 1979 CATACTTCTCAGCAGCATATCAAAATTTTACTGAAGGATTAAGTCATCTCTGGTGGTCTCAA 2038  
Db 1924 TATATGTTAACACCTTATGAATTTTATTAAGGATTAAGGCATCTCTGGAGGCTTAAA 1983  
QY 2039 TTGGCATTCATAGGAGTTGATGAAGCTCATCGTTTAAAAAATGATGACTCTCTCTGTGA 2098  
Db 1984 TTGGCATTTATAGGTTGATGAAGCACACCGGATTAAGAAATGATGACTCTCTGTGA 2043  
QY 2099 CAGGACTTTAATAGACTTTAAGTCCAAACCATGACTTCTGATTAAGTGAACCCCATGCA 2158  
Db 2044 TAAAACTTTAATAGATTTTAAATCCAATCATCTGCTCTCTTATCACTGGAACTCTCTACA 2103  
QY 2159 AAATTCCTCAAGAGCTGTGGTCTTGTGTCATTTTCATCATGATCCAGAAAAATTTTCCCTC 2218  
Db 2104 GAATTCCTCAAGAGCTGTGGTCTTGTGTCATTTTCAATTTATGCCAGAAAGTTTCTTC 2163  
QY 2219 CTGGGAAGATTTTGAAGAGGAGCATGGCAAGGAGAGAGATGGTTATGCAAGTCTTCA 2278  
Db 2164 CTGGGAAGATTTTGAAGAAGAACTGCAAGGAGGAGAAATATGGTTATGCAAGCCTTCA 2223  
QY 2279 CAAGAGCTTGACCACTTTTACTAAGAGAGTTAAAGAGATGAGAAAGTCTTTACC 2338  
Db 2224 CAAGGAGCTTGAGCCATTTCTGTGTACGCCGAGTTAAGAAAGATGTGGAATAATCTCTCC 2283  
QY 2339 TGCTAAGGTTGAACAAATCTGAGGATGGAATGAGTGCATTCAGAAGCAATATTACAA 2398  
Db 2284 TGCCAGGTTGACGAGATTTAAGATGGAATGAGTGTCTTACAGAAACAATATTACAA 2343  
QY 2399 GTGGATTTAAAGGAATTAAGCCCTCAGTAAAGGTTCAAAAGGCAAGTACCTCAGG 2458  
Db 2344 ATGGATTTTAACTAGGAATTTACAAAGCCCTCAGCAAGGTTTCCAAAGGCAAGTACCTCAGG 2403  
QY 2459 CTTTCTGAACATTTATGATGAACCTTAAGAGTCTGTGAACCATTTGCTACCTCATTAAGCC 2518  
Db 2404 CTTTTGAACATTTATGATGAGCTTAAGAAATGTTGAACCATTTGCTACCTCATTAAGCC 2463  
QY 2519 ACCAGATGATTAATGAATTTCTATAATAACAGAGGCTTTACAGCATTTGATACGTAGCAG 2578  
Db 2464 ACCAGATAATAATGAATTTCTATAATAACAGAGGCTTTACACACTTAATTCGTAGTAG 2523  
QY 2579 CGGAAACTAATCCTTCTTGACAAGCTACTGATTTGCTGTGCGAGACAGTGGGCAACAGAGT 2638  
Db 2524 CGGAAATTTGATTTCTTTGACAAGCTATTAATTTGCGCTAAGAGAACGAGGCAATCGAGT 2583  
QY 2639 TCTGATTTTCTCTCAGATGTTGAGGATGCTGGACATCTCTAGCAGAAATATCTGAAGTATCG 2698  
Db 2584 TCTATTTTTCACAAATGTTGCGGATGTTAGATATACTTGCGAATATTTGAAATATCG 2643  
QY 2699 CCAAGTTTCCCTCCAGAGACTTGTGATCAATTAAGGGAATTTGAGGAAGCAAGCACT 2758  
Db 2644 TCAATTTCCCTTTCAAAGATTAGATGATCAATTAAGGAGAACTGAGGAAACAAGCTCT 2703  
QY 2759 GGATCATTTCAATGCAGAGGATCAGAGGATTTCTGTTTTTACTGCTCTACAGAGCTGG 2818

Db 2704 AGATCATTTTAAATGCTGAGGATCAGAGGATTTTGTCTTTTGTCTCCACAAGAGCTGG 2763  
QY 2819 AGGATTAGGTATTAACCTTGGCATCTGCTGACACTGTAGTTATTTTTCATCTGACTGAA 2878  
Db 2764 AGGCTTAGGGATTAATTTTAGCCTCTGCTGACACTGTGTATATTTGATTTCCGATGGAA 2823  
QY 2879 TCCACAGAATGATCTGCAGGCACAGGCGAGAGCTCATAGAAATTTGGACAGAGAAACAGT 2938  
Db 2824 TCCACAGAATGATCTTCAGGCACAGGCTAGGCCATTCGAATTTGGCCCAAGAAACAGT 2883  
QY 2939 TAATATTTATCGCTAGTCTACAAAGGATCAGTAGAAGAAGATATTTCTTGAAGAGCCAA 2998  
Db 2884 GAATATTTATCGTCTAGTTTACAAAGGATCAGTTGAAGAAGATATTTCTTGAAGGCGCA 2943  
QY 2999 GAAGAGATGGTCTGATAGACCATTTAGTAATTCAGAGAAATGGACACAGAGAAACACTGT 3058  
Db 2944 AAAGAAGATGGTTTTAGATCATCTGTAAATCAAGAAGATGGACACAACCTGGGAGACAGT 3003  
QY 3059 TCTGCAATACAGGTTCAACTCCATCAAGCTCTACACCTTTTAAATAAAGAAGATTATCAGC 3118  
Db 3004 ACTACATACAGGTTTCCGCCATCAAGTTCTACTCTTCAATTAAGAAGAGTTATCAGC 3063  
QY 3119 TATTTTGAAGTTTGGTCTGAGGAACCTTTTAAAGAACCTGAAAGGAGAAACAAGAGCC 3178  
Db 3064 CATTTTAAAGTTTGGTCTGCAAGAACTTTTAAAGAACCTGAAAGGAGAAACAAGAGCC 3123  
QY 3179 CCAGGAATGGATATAGATGAATACTTGAAGAGAGCTGAAACTCGGGAATAAGGCCAGG 3238  
Db 3124 CCAGGAATGGATATAGATGAATACTTGAAGAGAGCTGAAACTCATGAAATGAACCCAGG 3183  
QY 3239 TCCATTTGACTGTAGGGATGAGTTGCTTTTCAAGTTTCAAGCTGGCGAACTTTTCCAAATAT 3298  
Db 3184 TCCTTTAACTGTAGGAGATGAATTTGCTTTCCAGTTTCAAGTTTCCAACTTCTCAATAT 3243  
QY 3299 GGATGAAGATGATATTAGTTTGAACCCAGAAAGAAATTTCAAGAAATTTGGAAGAAATCAT 3358  
Db 3244 GGATGAGGATGACATTTAGTTTGAACCTGAAAGAAATTTCAAGAAATTTGGGAGGAAATAT 3303  
QY 3359 CCAGGAATCCCAACGGAGAGATAGAGGAGGAGGAGAAAGACAAAGAACTTGAAGAAAT 3418  
Db 3304 TCCAGAGATCAAGAAAGACGATTTAGAAAGAAAGAAAGAAAGAACTTGAAGAAAT 3363  
QY 3419 ATACATGCTCCCGAGGATGAGAACTGTGCAAAACAGATCAGCTTTAATGGGAGTGAAG 3478  
Db 3364 TTATATGCTCCCAAGATGAGAAATTTGCAAAACAGATTTAGTTTCAATGGAAGTGAAG 3423  
QY 3479 AAGACGAGTAGGAGCAGAGATATTCTGGATCTGTATGACTCCATCAGAGAAAGAA 3538  
Db 3424 GAGGCGCAGTAGAAGTAGGAGATCTCTGGATCTGTATGATTTCCATCTCAGAAAGGAA 3483  
QY 3539 ACGGCCAAAAACGCTGGAGACCTCGAACCTTCTCGAGAAATATTAAAGGATTTAG 3598  
Db 3484 AAGCCAAAGAAACGTTGGAAGACCGGACTATTTCTCGGAGAAATATTAAAGGATTTAG 3543  
QY 3599 TGATGACAGATCAGCGGTTTATCAAGAGTTTACAAGAAATTTGGTGGCCCTCTGGAAG 3658  
Db 3544 TGATGACAGAAATTTAGCGGTTTATCAAGACCTATAAGAAATTTGGTGGTCTCTGGAAG 3603  
QY 3659 GTTAGATGCTGTAGCTAGATGCTGAATGGTTGATTAATCTGAGACAGACCTTAGACG 3718  
Db 3604 ATTAGATGCAATTTGCTCGAGATCTGAGTTAGTTAGTAACTCAGAAACAGACCTTAGACG 3663  
QY 3719 TTTGGGTGAATTTGATCAATAATGATGATTAAGGCTTTTAAAGGACAAATTTCACTCGACA 3778  
Db 3664 ACTGGAGAAATTTGATCAATAATGTTGCAATTAAGCAATTAAGGATAGTTTCTTCAGGAAC 3723  
QY 3779 AGAAGACGAGGAGGTAGACTTTGGGAAAGTTAAAGGCCCAACGTTTCTGAATCTCAGGAGT 3838  
Db 3724 AGAACGAACAGGTGTAGACTCGGAAAGTGAAGGTTCCCAACATTTCCGAATATCAGGAGT 3783  
QY 3839 GCAGGTGAATGCAAAACACTAGTCTATCTCAGGAAGAGAGCTGGCCACTGCAACAATC 3898  
Db 3839 GCAGGTGAATGCAAAACACTAGTCTATCTCAGGAAGAGAGCTGGCCACTGCAACAATC 3898





RESULT 4  
MUSCHD1X  
LOCUS MUSCHD1X 5349 bp mRNA linear ROD 19-SEP-1996  
DEFINITION Mouse DNA-binding protein (CHD-1) mRNA, complete cds.  
ACCESSION L10410 X66028  
VERSION L10410.1 GI:455014  
KEYWORDS DNA binding protein.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 5349)  
Delmas,V., Stokes,D.G. and Perry,R.P.  
A mammalian DNA-binding protein that contains a chromodomain and an  
SNF2/SWI2-like helicase domain  
Proc. Natl. Acad. Sci. U.S.A. 90 (6), 2414-2418 (1993)  
93211972  
2 (bases 1 to 5349)  
Perry,R.P.  
Direct Submission  
Submitted (08-APR-1993) Robert Perry, The Fox Chase Cancer Center,  
7701 Burholme Avenue, Philadelphia, PA 19111, USA  
On Feb 16, 1994 this sequence version replaced gi:293322.  
LOCATION/Qualifiers  
1..5349  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/cell\_type="plasmacytoma"  
1..5349  
/gene="CHD-1"  
171..5306  
/gene="CHD-1"  
/codon\_start=1  
/product="DNA-binding protein"  
/protein\_id="AAB08486.1"  
/db\_xref="GI:455015"  
/translation="MNGHSDESIVRNGSGESSQSDGCGSAGSGSGSSSSGSSDSS  
SOGSSDSSDSSGSSSESDETSRENKVAQPKPKVDGAEFWKSPSITLAVORAML  
RKPOAQOQRPASSNGSSESDSSDSSGAKRKNHNDWOMSGSPSOLG  
SDSESEEDKSSCDGTESDYEPKNKRSKPQNRKSKNGKILGOKRKQIDSSDE  
DDEDYDNKRSRRQATVNSYKEDDEMKTDSDLEVCGEDVPDPEEFETIERV  
DCRVGRKATGATTTIYAEADGDPNAGFERNKPEGDIOYLKRWGSHIHNTWTEE  
TLKQNRGVMKLDNYKKKQDKRWKNASPEDEYNNCOELDDHKOYIYERI  
IAHSNQSAAGLPDYCKWQGLPYSECSWEDGALISKFKQTCIDYFNRNOSKQTPFK  
DCKVLKQRPVALKKQPSYIGGHEGLELRYOLNGLNLWLAHSCWKGNSCILADEMGL  
CKTIOTISFLNVLFEHOLYGFLLVPLSLTSHQRELOTWASOMNAVYVGLDINSR  
NMRTHWHWQTKRLKLNILLTYEILLKKAFLGLNWFIFIGVDEAHRUKNDLSLL  
YKTLIDFKSNHRLITGTPQLNSKELMSLHFMPEKPSWEDFEHGHGREYGYA  
SLHKELEPFLLRVKKDEKSLPAKVEQILRMESALQYKWLITRYKALSGSK  
GSTGFLIMMELKCCNHCYLIKPPDNNEFYKQELHILIRSGKILILDLKLIRL  
RERGNRLIFSOVMRLDILAEYLYKROFPFORLDGSIKGLRKOALDHFNAEGSEDF  
CFLSTRAGGLGINLASADTVIFDSNDPNQDLOAQAHRIGOKKOVNIYRLVTKG  
SVEEDILERAKKWLHLVLTQMDTCTKVLHTGSAPSSSTPPNKBELSAILKFGAE  
ELFKPEGEQEPQEMDDELKRAETHENEPGLUSVGDLLSPKVANFNMDEDDI  
LELPNRKNWEEIIPQORRLKEEROKDELEYMLPRMNCAKQISFNGSEGRS  
LSRRYSGSDSDSISRRKPKRGRTPIPRENIGFSDAEIRRFKSKYKFGPERL  
DAIARDLSDKSDTLRLGLVHNGCVKALKDSSGTERAGRLGKVPFTRISG  
VOVNAKLVIATHEDELIPLHKSIPSPDEERKQYTPCHTKAHDIDMCKEDSDNLLIG  
IYEYCGSWEMIKMPDLSLHKILPDDEKPKQAKQLQTRADYLIKLLSRDLAKREA  
QIRCGAGSKRRKTRAKSKAMKSIVKEETKSDSPSPSEKSDDEDDKLDSPESK  
DRSKSVSDAPVHTASGEPVPIAESEELDQKTFISCKERMREVKAALKQLDRPEK  
GLSREQVLEHTRQCIKIGDHITECLKEYSNPEQIKQWRKNLWIFVSFTEFDARKLH  
KLYHAIKROHSDSYKSDSRKRPYSFNSGCKHREMDHYQDSDSYSDREKHKRLDD  
HRSRHRPSLEGGLKDRCHSDHRSMDHRSSEHSHHKSRRDYRLYLSWQOLD  
HRAASSGPRSPLDQSPSGRSRPFPHSAHRSTPETHWSSRKT"  
BASE COUNT 1739 a 1067 c 1319 g 1224 t  
Query Match 51.4%; Score 3398.2; DB 10; Length 5349;  
Best Local Similarity 79.6%; Pred. No. 0;

	Matches 4150;	Conservative	0;	Mismatches 1008;	Indels	57;	Gaps	9;
Qy	201	TTTTCTTTTAACTTCTTAATCTTACAATGAATGGGCACAGTCATGAGAAAGCTGTGAAGA						
Db	144	TTTTATTAAACACTTCTTGATATTATTAATGAATGGACACAGTCATGAAGAAAGCTGTGA						
Qy	261	AACAGCAGTGGAGAGTCAAGCAGATCAGATGATGATTTCTGGGTGAGCTTCAGCTTCTGGA						
Db	204	AATGGCAGCGAGAAATCAAGTCAGTCAGGTGATGATTTGGGTGAGCTTCAGCTCTGGA						
Qy	321	TCGTGTTCAAGCTCTGGAAGCAGTAGCGATGGAAGTAGCAGCCAGTCAGTAGCAGTAC						
Db	264	TCGTGGTCGAGTTCTGCGACGACAGTCAGCGAAGCAGCAATCCGGAGGACGCGAC						
Qy	381	TCGTGATCTGGTTTCAGAGTCAGCAGTCATCCGAATCAGAGTCGACACATCTAGAGAG						
Db	324	TCGTGATCTGGCTCTGACTCAGGAAGTCAATCAGAGTCGTAATCAGACACATCCCGAGAG						
Qy	441	AAGAAACAAAGTTCAAGCTAAACCTCCGAAAGCTGACGGATCTCGAGTTTGGAAAGTCCAGT						
Db	384	AACA---AGTTTCAAGCAAAACCCACCACCAAGTCGACGGAGCCGAGTTTGGAAATCTAGC						
Qy	501	CCAGCATCTACTGCTGTACAGATCAGCAGTCTCTCAAGAGCAAC-----AGCAA						
Db	441	CCAGTATTCTGGCTGTCCAGAGATCTGCAATGCTTAGGAAGAGCCACAGAGGCCAG						
Qy	552	CAGCAAAAAGCAGCATCATCAGACAGTGGTTCAAGAAGAGGACTCATCCAGTACTGAAGAT						
Db	501	CAGCAGCGCCAGCTTCATCTAATAGTGGATCCGAAGAGACTCTGTCACGACGTGAAGAC						
Qy	612	TCGCCGATGATTTCGTCAGTGAACCTAAGAAAGAAAGCATAAAGATGAAGACTGGCAA						
Db	561	T---CCGACGACTCGTCCAGCGGTGCCAAGAGGAAGACACAATGATGAAGACTGSCAG						
Qy	672	ATGTCAGGTCAGGTCAGTATCAGGAAGTGGTTCTGATTTCTGAATCGGCGGAGATGGG						
Db	618	ATGTCGGTCCGGATCTCCATCTCAGCTCGGTTCTGAGACTCAGAATCTGGAAGAGAGCGA						
Qy	732	GATAAAGCAGTTGTGAAGAAAGTGAATCTGACTATGAGCAAAAACAAAGTCAAAAGC						
Db	678	GATAAAGCAGCTGCGACGGGACAGATCCGACTACGACCGCAAAAACAAAGTCAAGAAC						
Qy	792	CGTAAACCTCCAAGCAAGTAATAAGCCAAAAGTGGGAAAAGAGCAGCAGAGAGAAG						
Db	738	CGAAGCCCTCAGAATAGATCTAAGTCAAAAATGGGAAAAAATTTCTTGACACAAAAAG						
Qy	852	AGGCAACTTGATTCATCAGAGGAGGAGGAGGATGATGAAGATTATGATAAGAGGA						
Db	798	AGACAGATTGATTCATCTGAGGATGAAGATGATGAAGATTATGATAAACGAAGC						
Qy	912	TCCTGTCGCGCAGCAACAGTGAATGTTAGTTACAAAAGAGCTCAAGAAACCAAGACAGAT						
Db	858	TCTCGCGCAACCCACCGTCAATGTGAGCTACAGGAGGATGAAGAAATGAAAACCTGAC						
Qy	972	TCGTGATGTTGCTGGAAGTTTCTGGAGAGGATGTCCCACAGACTGAAGAAGATGAATTT						
Db	918	TCGGATGACTGCTGGAGGTCTCGCGGAGGAGCTCCCTCAGCTGAGGACGAGGAGTTT						
Qy	1032	GAACTATAGAGAAAGTTTATGGACATCGAATTTGGCGGAAAAGAGGCCACTGGTGCCTCA						
Db	978	GAGACAAATAGAGAGGTTTATGGATTTCAGATGGGCGGAAAAGGAGCTACTGTTGTTACT						
Qy	1092	ACCACCATCTATGCCCTTGAGGCAGATGTTGACCCCAATGCTGGGTTTGAAGAGTCAAG						
Db	1038	ACAACCATTTATGCTGTCGAGACAGATGGTGACCCCAATGTCAGGATTTGAAAGAACAAA						
Qy	1152	GAGCTGGGAGAAATACAGATATCTTTAATGGAAGGCTGGTGCACACATCCACATAACACT						
Db	1098	GAGCCAGGAGACATACAGATATTTAATTAAGTGGAAAGGATGGTCTCACATCCCAACACA						
Qy	1212	TGGGAACACTGAAGAAACGCTGAAGCAACAAATGTTTAAAGGAATGAACAACCTGGCAAC						
Db	1158	TGGGAGACAGAGAACCCCTGAAGCAGACAGACGTTAGAGGGATGAAAAAATTTGGATAAT						

QY 1272 TACAAGAAAAGGATCAGAGACAAAACGCTGGCTGAAAAATGCTTCTCCAGAGATGTG 1331  
DB 1218 TATAGAAAAAGATCAAGAGACGAACGATGGCTGAAAAATGCTTCTCCAGAGATGTG 1277  
QY 1332 GAATATTATTAACCTGCGCAGGAGCTTACAGATGATCTGCAACAAATATCAAAATAGTG 1391  
DB 1278 GAATATTATTAATGGCAGCAAGAGCTTACAGATGATCTACACAAACAGTATCAGATAGTG 1337  
QY 1392 GAAAGAAATTAATGCTCATTTCAATCAAAAGTCCAGCAGCTGGTTATCCGAGCTACTATTGC 1451  
DB 1338 GAGGCATAATTTGCTCATTTCCATCAAAAATCAGCAGCTGGTCTCTGATTAATATTGC 1397  
QY 1452 AAATGGCAGGGCTGGCTTACTCAGATGATGAGTATTTAGCAGAAATCAATCCAGACTACTCCCTTT 1571  
DB 1398 AAATGGCAGGGCTTCCATCTCAGATGCTACCTGGGAGATGGAGCTCTCATTTCCAAA 1457  
QY 1512 AAGTTTCAGCAGGATGATGATGATTTTACAGAAATCAATCCAGACTACTCCCTTT 1571  
DB 1458 AAGTTTCAGATGATGATGATTTTACAGAAATCAATCCAGACTACTCCCTTT 1517  
QY 1572 AAGGACTGAAGGCTTAAACAGAGACCAAGTTTGTGATGATGAGTGAAGAACCAACATCT 1631  
DB 1518 AAGGTTGCAAGTGTGTAACAAAGACCAAGATTTGTGATGATGAGTGAAGAACCAACATCT 1577  
QY 1632 TACATTTGGAGCATGAAAGTCTGGAGTTAAGAGATTAATCAGTTAAATGGATTTGAATGG 1691  
DB 1578 TATATTGGAGCATGAGGGCTTAGAAGTGAAGATTAATCAGTTAAATGGATTTGAATGG 1637  
QY 1692 CTCGCTCATTTGATGCAAGGAAATAGTTGATTTCTGAGATGAAATGGTCTGGGT 1751  
DB 1638 CTCGCTCACTCTTGTGCAAGGAAATAGTTGATTTCTGAGATGAAATGGTCTGGGT 1697  
QY 1752 AAACAATACAAACATTTCTTTCTGAACTACCTGTTTCAATCAACATCACTGATGGC 1811  
DB 1698 AAACAATACAGAGATCTCATTTTGAAGTATTTGTTCCATGAACATCACTGATGGG 1757  
QY 1812 CTTTCTCTCGGGTGCACCTTCTGAGTACCTGAGTCTGCAAGTCTGCAAGAGATTTCAACT 1871  
DB 1758 CTTTCTCTAGTTGCTCCGCTCTCCACATGATCTTCTGCGAGGAGATTCAGCG 1817  
QY 1872 TGGGCTCTCAGATGAATGCTGTGATTTACTTAGAGATTAATCACTAGTAAATATGATA 1931  
DB 1818 TGGGCTCTCAGATGAATGCTGTGATTTACTTAGCGGACATTAACAGCAAAACATGATA 1877  
QY 1932 AGGACTCATGAATGATGATCCACAGACTAAACGATTAAGTTTAAACATCTCTGAGC 1991  
DB 1878 AGAAGTATGAATGATGATCCACAGACCAACCGGTTAAATTTAATATATCTTTAACA 1937  
QY 1992 ACATATGAAATTTTACTGAAGGATTAAGTCAATCTTGTGCTGCTCAATTTGGGCAATTCATA 2051  
DB 1938 AGGTATGAAATTTTATTTGAAGGATTAAGGATTTCTTGTGCTGCTCAATTTGGGCAATTCATA 1997  
QY 2052 GGAGTTGATGAAGCTCATCTTTTAAATAATGATGATCTCTCTGACAGGATTTTAAATA 2111  
DB 1998 GGTGTTGATGAAGCGCATCGATTAAGATGATGATTTCCCTCTCTGACAAACTTTAATC 2057  
QY 2112 GACTTTAAGTCAACCATGCTGATTTGATTAAGTCAACCTGCAAAATTTCCCTCAAA 2171  
DB 2058 GACTTTAATCTAACCATGCTCTGATCACTGGAACCTCTACAGAACTCTCCCTGAG 2117  
QY 2172 GAGCTGTGCTTTGTTGCTCATTTCAATCATGCGCAAAATTTTCTCTGGAAGATTTT 2231  
DB 2118 GAGCTGTGCTGCTGCTCTTTGCTGAGGAGGAGTTTCTTCAATGGAAGATTTT 2177  
QY 2232 GAAGAGCATGGCAAGAGAGATGATGTTATGAGTCTTCAAAAGCTTCTCAAAAGCTTGAA 2291  
DB 2178 GAAGAGCATGGCAAGAGAGATGATGTTATGAGGCTTCCACAGAGCTTGAG 2237  
QY 2292 CCATTTTACTAAGAGAGTTTAAAGAGATGAGAAAGTCTTTTACCTGCTAAGGTTGAA 2351  
DB 2238 CCATTTCTGTAGCAGAGTTAAGAAAGATGAGAAATCTCTCTCCGCAAGGTTGAG 2297

QY 2352 CAAATTTCTGAGGATGGAATGAGTGCATTCAGAGCAATATTTACAAGTGGATTTAACA 2411  
DB 2298 CAGATTTTAAAGATGAGATGAGTGCCTTTACAGAGCAATATTTACAAGTGGATTTAACA 2357  
QY 2412 AGGAATTTAAAGCCCTCAGTAAAGGTTTCAAAAGGAGTCTACCTCAGGCTTTTCTGAACAT 2471  
DB 2358 AGGAATTTAAAGCCCTCAGCAAAAGGTTTCCAAAGGAGTCTACCTCAGGCTTTTGAACAT 2417  
QY 2472 ATGATGGAATCTTAAGAGTGTGTTAAACATTTGCTACCTCATTAAAGCCACAGATTAAT 2531  
DB 2418 ATGATGGAATCTTAAGAGTGTGTTAAACATTTGCTACCTCATTAAACCCACAGATTAAT 2477  
QY 2532 GAATTTCTATTAACACAGAGGCTTACAGCATTTGATACCTAGCAGGAGGAACTAATC 2591  
DB 2478 GAATTTCTATTAACACAGAGGCTTACACACTTAATCCGTAGTACGGAACACTCATC 2537  
QY 2592 CTTCTTGACAGCTACTGATTTGCTGCGAGAGCTGGCAACAGCTTCTGATTTCTCT 2651  
DB 2538 CTTCTGCAAGCTGTTGATTCGCTAAGAGAAAGAGCAACCGTGTCTATTTCTCT 2597  
QY 2652 CAGATGAGGATGCTGGACATCTAGCAGAAATATCTGAAGTATCCGAGTTTCCCTTC 2711  
DB 2598 CAGATGAGGATGTTAGACATATCTGCGAGATTTTGAAGTACCGCAGTTTCCCTTT 2657  
QY 2712 CAGAGATTTGATGGATCAATAAAGGGGAATTTGAGGAAGCAAGCACTTGGATTTCAAT 2771  
DB 2658 CAAAGATTTAGATGATCGATATAAAGGGAGCTGAGGAAGCAGGCTTTGGATCACTTAAT 2717  
QY 2772 GCAGAGGATCAGAGGATTTCTGTTTTTACTGCTACAAAGAGCTGGAGGATTTAGGTAT 2831  
DB 2718 GCCGAGGCTCAGAGGATTTCTGCTTTCTACCGAGCTGGGGTTTAGGCAAT 2777  
QY 2832 AACTTGGATCTCTCAGACTGATTTATTTTGTCTGCTGGAATCCACAGATGAT 2891  
DB 2778 AACCTAGCTCTCTCAGACCTGTTTATATTCGATTTGATTCGAATCCACAAATGAT 2837  
QY 2892 CTGAGCAGCAGGAGAGCTCATAGATTTGACAGAGAAACAGCTTATATTTATCGG 2951  
DB 2838 CTTAGGCTCAGGAGAGCTCCGATTTGACAGAGAAACAGCTTATATTTATTCG 2897  
QY 2952 CTAGTCACAAAGGATCAGTAGAGAGATATTTTGAAGAGCCAGAGAGATGGTG 3011  
DB 2898 TTGGTTTAAAGGATGATGATGAGAGATTTCTTGAAGGGGCAAAAGAAATGTT 2957  
QY 3012 CTAGACATTTAGTAATCTAGAGATGACACAGAGAAACCTGTCTGCAATACAGGT 3071  
DB 2958 TTGGATCATCTGTTTATTCAAAGATGATACCACTGGGAAGACAGTGTGACACAGGC 3017  
QY 3072 TCAACTCCATCAGCTCTACACCTTTTAAAGAGAGATTTATCAGCTATTTTGAAGTT 3131  
DB 3018 TCGGCTCGCTCAAGTTCCACCTTCAATTAAGAGGATTTATCCGCAATTTTAAATTC 3077  
QY 3132 GGTGCTGAGGAACTTTTAAAGACCTGAAGAGAGAAACAGGAGCCAGGAAATGGAT 3191  
DB 3078 GGTGCTGAGGAGCTTTTAAAGAGCTGAAGAGAGAGAGAGGAGGAGGAGGAGGAT 3137  
QY 3192 ATGATGAAATTTGAAGAGCTGAACTCGGAAATAGAGGAGGAGGAGGAGGAGGAGGAG 3251  
DB 3138 ATGATGAAATTTGAAG 3197  
QY 3252 GGGGATGATGCTTTTCAAGTTCAAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAG 3311  
DB 3198 GGGGATGAGCTGCTTCTCAGTTTARGTTTARGTTTARGTTTARGTTTARGTTTARGTTTARG 3257  
QY 3312 ATTGATTTGGAACCAAGAAAGAAATTTCAAGAAATTTGGGAAGAAATTCATCCAGAAATCCCA 3371  
DB 3258 ATTGATTTGGAACCTGAAAGAAATTTCAAGAACTGGGAGAGATTCATTCAGAGAGAGCA 3317  
QY 3372 CGGAGAGGATTAGAGGAGGAGAAAGAGCAAAAGAACTTTGAAGAAATATATCATGCTCCG 3431  
DB 3318 AGACGCGACTTAGAGAGGAGGAGAGAGCAAAAGAACTGGAGGAAATTTATATGCTCCG 3377  
QY 3432 AGGATGAGAAATCTGCAAAAACAGATCAGCTTTTAAAGAGAGTGAAGAGAGAGAGAGAG 3491



## AUTHORS

Fridolfsson, A. K. and Ellegren, H.

## TITLE

Molecular evolution of the avian CHD1 genes on the Z and W sex chromosomes

## JOURNAL

Genetics 155 (4), 1903-1912 (2000)

## MEDLINE

20384598

## REFERENCE

2 (bases 1 to 2943)

## AUTHORS

Fridolfsson, A.-K. and Ellegren, H.

## TITLE

Direct Submission

## JOURNAL

Submitted (30-AUG-1999) Department of Evolutionary Biology, Evolutionary Biology Centre, Uppsala University, Norbyvägen 18D, Uppsala S-75236, Sweden

## FEATURES

Location/Qualifiers

source

1..2943

/organism="Gallus gallus"

/db\_xref="taxon:9031"

/chromosome="W"

&lt;1..&gt;2943

/gene="CHD1"

&lt;1..&gt;2943

/gene="CHD1"

/codon\_start=1

/product="chromodomain helicase DNA binding protein 1"

/protein\_id="AAD56023.1"

/db\_xref="GI:5917755"

/translation="IQYLIKRWKWSHINHTWESEETLKQNVKMKKLDNYKKCOBT

KRWLNKASPEDVEYVNCQEFDDHKKQYVVERIASHNOKSAAGPYDYCKWQGLP

YSEKSWEDGALIAKFAQIDEYFSRNSQKTPFKDKILKQRFVALKPKPSYIGG

RDLLELDYQNLGNLWLAHCKNGSCILLADEMLGKT IOT ISFLNLYLPHHELYGPF

LLVVPFLSTLSWOREIOTWAPQNVAVYLDITSRNMLRTHEMMHPTKRLKFNLLT

TYEILLKDKSFLGLNRVFIIGIDEAHLKNDSDLSLYKTLIDFKSNHRLITGTPLNT

LKEWLSLLHFIMPEKFSWEDFEHKGKRGFYASLHKEFLPLLRKVKDVEKSSP

AKVEQILRMEMSAKQYKWLIRNYKALSKSGKSGFLNMLKMKCNHCYLI

KPPDNEFYNKQALQHLIRSSGKLLLDKLLIRLRERGNRLVPSQVMDLILAY

LKYPQFQRLDGSIKGLRQALDHNAGSEDFELSTRAGGLINLASADYVI

PDSMDNPDLQAQRAHRIQPKKQVNIYLVTKGSVEEDILBRKAKMVLHVIOR

MDYTKVLTGSPSSSTPKNKEELSAILLKCAEELFKPEESEEPEPQMDIDELK

RAVRENESGLIYVAGDELLSQFYANFSNMDEDDILEPQNLKNEELIPEYQWRI

EEERQKLEIYVMPRMNCAKQISFNQNEGRCSRRYSRSGSDSDISERKPKKKG

RFTPIIRNKGFSDAEIRFKISFKFQGLERLDAIARDAELVDKSDIDRLKGLP

VHNGCICALNDPFGQGRFGKVGFTFRIGVQVNAKLVISSHEELAPLHKRSL

SPERKRYVPIYHTKAAHFDDWGRKEDDSNLLIGIYEGYGSWEMIKMDPDLSTOK

ILPDDPPKK"

BASE COUNT 1023 a 478 c 663 g 779 t

## ORIGIN

Query Match 38.4%; Score 2538.2; DB 5; Length 2943;

Best Local Similarity 91.4%; Pred. No. 0;

Matches 2690; Conservative 0; Mismatches 253; Indels 0; Gaps 0;

Qy 1164 ATACAGTATCTTATTAATGGAAGGCTGTCACACATCCATTAACACTTGGGAAACTGAA 1223

Db 1 ATACAGTATCTTATTAATGGAAGGCTGTCACACATTCATTAATCTTGGGAAGTGAA 60

Qy 1224 GAAAGCGTGAAGCAACAAATGTTAAAGGAATGAACAACTGGCAACTACACAAAGAAAG 1283

Db 61 GAAACTCTGAAGCAACAAACGTTAAAGGAATGAAGAACTAGACAAATTAAGAAAG 120

Qy 1284 GATCAGAGACAAACGCTGCTGAAAATGCTCTCCAGAACTGTGGAATATTATAAC 1343

Db 121 TGTCCAGCAACAAACGATGCTGAAAATGCTCTCCAGAACTGTGGAATATTATAAC 180

Qy 1344 TCCAGCAGAGCTTACAGATGATCTGCACAAACAAATATCAATAGTGAAGAAATTAAT 1403

Db 181 TCCACACAGATTTACAGATGATCTGCACAAACAAATATCAATAGTGAAGAAATTAAT 240

Qy 1404 GCTCATTTCAAACTACAGAGCTGTTATCCGAGTACTATTGCAAAATGCGAGG 1463

Db 241 GCTCATTTCAAACTACAGAGCTGTTATCCGAGTACTATTGCAAAATGCGAGG 300

Qy 1464 CTGCTTACTCAGATGTAGTGGGAAGATGGTCTCTCTTCCAAAAGTTTCAGGCA 1523

Db 301 CTGCTTACTCAGATGTAGTGGGAAGATGGTCTCTCTTCCAAAAGTTTCAGGCA 360

Qy 1524 CGCATGTGATGAGTATTTTACGAGAAATCAATCCAAAGACTACTCCCTTTAAGGACTGCAAG 1583  
Db 361 TGTATTGATGAGTATTTTACGAGAAATCAATCCAAAGACTACTCCCTTTAAGGACTGCAAG 420  
Qy 1584 GTTCTAAACAGAGAGCAAGATTTTGTGCACACTGAAGAAGCAACATCTTACATTTGAGGA 1643  
Db 421 ATTCTAAACAGAGGCGCAAGTTTGTGCGCTTAAAGAAGCAACATCTTATATTGAGGA 480  
Qy 1644 CATGAAGCTGTCAGATTTAGAGATATACGTTAAATGGATTTGAATGGCTGCTCATTTCA 1703  
Db 481 CCGGATTTGTTAGAGTTAAGAGATATACGTTAAATGGATTTGAATGGCTGCTCATTTCA 540  
Qy 1704 TGGTCAAGGAATAGTTGTTATCTTTCAGAGATGAATGGCTGCGGTGAACAAATACAA 1763  
Db 541 TGGTCAAGGAATAGTTGTTATCTTTCAGAGATGAATGGCTGCGGTGAACAAATACAA 600  
Qy 1764 ACAATTTCTTTCTGAACACTGCTTTTCATGAACATCAACTGTATGGCCCTTTCTCTG 1823  
Db 601 ACAATTTCTTTCTGAACACTGCTTTTCATGAACATCAACTGTATGGCCCTTTCTCTG 660  
Qy 1824 CCGGTGCACCTTTCTACCTTGGACATCTTGGCAAGAGAGATTCAAACTTGGGCTCCTCAG 1883  
Db 661 GTTGTGCCACTTTCTACCTTGGACATCTTGGCAAGAGAGATTCAAACTTGGGCTCCTCAG 720  
Qy 1884 ATGAATGCTGTAGTATTCTAGGAGATATAAATAGTAGAATAATATGATGAAGACTCATGAA 1943  
Db 721 ATGAATGCTGTAGTATTCTAGGAGATATAAATAGTAGAATAATATGATGAAGACTCATGAA 780  
Qy 1944 TGGATGCATCCACAGACTTAAACGATTAACGTTTAAACATATCTCTGACGACATATGAAAT 2003  
Db 781 TGGATGCATCCACAGACTTAAACGATTAACGTTTAAACATATCTCTGACGACATATGAAAT 840  
Qy 2004 TTACTGAAGATAGTCACTCTGCTGGTGTCTCAATGGGCAATTCATPAGGACTTCATGAA 2063  
Db 841 TTGCTGAAGATAGTCACTCTGCTGGTGTCTCAATGGGCAATTCATPAGGACTTCATGAA 900  
Qy 2064 GCTCATGCTTTAAAAATGATGACTCTCTCTGACAGACTTTTAATAGACTTTAAGTCC 2123  
Db 901 GCTCATGCTTTAAAAATGATGACTCTCTCTGACAGACTTTTAATAGACTTTAAGTCC 960  
Qy 2124 AACCATGACTTCTGATTTACTGGAACCCCACTGCAAAATCCCTCAAGAGCTGTGGTCT 2183  
Db 961 AATCATGCTTCTTATTTACTGGAACCCCACTGCAAAATCCCTCAAGAGCTTTGGTCT 1020  
Qy 2184 TTGTTGCATTTTCATCATGCCGAAATTTTCTCTCTGGAAGATTTTGAAGAGGAGCAT 2243  
Db 1021 TTGTTGCATTTTCATCATGCCGAGAAAGTTTCTCTCTGGAAGATTTTGAAGAGGAGCAT 1080  
Qy 2244 GGCAAGGAGAGAGATGATGTTTATGCAAGTCTTTCACAAAGAGCTTGAACCATTTTACTA 2303  
Db 1081 GGGAAAGGAGAGAGTTTGGTTTATGCAAGTCTTTCACAAAGAACTTTGAGCCATTTCTACTC 1140  
Qy 2304 AGAAGAGTTAAAAAGATCTGAAAAGTCTTTTACCTGCTTAAGTTGAACAAATTTCTGAGG 2363  
Db 1141 AGAAGAGTTAAAAAGATCTGAAAAGTCTTTTACCTGCTTAAGTTGAACAAATTTTAAAG 1200  
Qy 2364 ATGGAATAGTGCATTTGAGAGCAATATTACAGTGGATTTTAAACAGGAATTTATAA 2423  
Db 1201 ATGGAATAGTGCATTTGAGAGCAATATTACAGTGGATTTTAAACAGGAATTTATAA 1260  
Qy 2424 GGCCTCAGTAAAGGTTCAAAAGGCACTACCTCAGGCTTTCGACATTTATGGAAGCTT 2483  
Db 1261 GGCCTCAGTAAAGGTTTCAAGGCACTACCTCAGGCTTTCGACATTTATGGAAGCTT 1320  
Qy 2484 AGAAGTGTGTGAACCATTTGCTTACCTCATTAAGCACCAGATGATGAATGAATTTTATAT 2543  
Db 1321 AGAAGTGTGTGAACCATTTGCTTACCTCATTAAGCACCAGATGATGAATGAATTTTATAT 1380  
Qy 2544 AAACAGGAGGCTTTACAGCATTTGATACGTAGCAGCGGGAACCTTAATCTCTTTCACAG 2603  
Db 1381 AAACAGGAGGCTTTACAGCATTTGATACGTAGCAGCGGGAACCTTAATCTCTTTCACAG 1440

```
QY 2604 CTACTGATTCGTCCTGCGAGAACCTGGCAACAGAGTCTTCGATTTTCTCTCAGATGGTGAGG 2663
Db 1441 TTACTGATTCGTCCTACGAGAACCTGGCAACAGAGTACTGATTTTCTCTCAGATGGTGAGG 1500
QY 2664 ATGCTGACATCTCAGCAGATATCTGGAAGTATCGCCAGTTTCCCTTCAGAGACTTGAT 2723
Db 1501 ATGCTAGACATCTCAGCAGATATCTGGAAGTATCGCCAGTTTCCCTTCAGAGACTTGAT 1560
QY 2724 GGATCAATAAAGGGGAATTGAGCAAGCAGCAGTGGATCATTTTCAATGCAGAGGATCA 2783
Db 1561 GGATCAATAAAGGGGAATTGAGCAAGCAGCAGTGGATCATTTTCAATGCAGAGGATCA 1620
QY 2784 GAGGATTTCTGTTTTTACTGCTACAGAGCTGGAGGATTAGGTATTAACCTTGGCATCT 2843
Db 1621 GAGGATTTCTGTTTTTACTGCTACAGAGCTGGAGGATTAGGTATTAACCTTGGCATCT 1680
QY 2844 GCTGACACTGTAGTTATTTTGATTTCTGACTGGAATCCACAGATATCTGCAAGGACAG 2903
Db 1681 GCTGACACTGTAGTTATTTTGATTTCTGACTGGAATCCACAGATATCTGCAAGGACAG 1740
QY 2904 GCGAGAGCTCATAGAAATGGACAGAAAGAACAGGTTAATTTATCGGCTAGTCACAAA 2963
Db 1741 GCTAGAGCTCATAGAAATGGACAGAAAGAACAGGTTAATTTATCGGCTAGTCACAAA 1800
QY 2964 GGATCAGTGAAGAAGATATCTTTGAAGAGCCAGAAAGAGATGGTGTAGACCAATTA 3023
Db 1801 GGATCAGTGAAGAAGATATCTTTGAAGAGCCAGAAAGAGATGGTGTAGATCATTTA 1860
QY 3024 GTAATTCAGAGATGNCACGACAGAAACCTGTTCTGCATACAGGTTCACTCCATCA 3083
Db 1861 GTGATTCAGAGATGNCACGACAGAAACCTGTTCTGCATACAGGTTCACTCCATCA 1920
QY 3084 AGCTCTACACCTTTTAAAGAAAGAGTTATCAGCTATTTTGAAGTTTGGTGTGAGGAA 3143
Db 1921 AGCTCANCACCTTTTAAAGAAAGAGTTATCAGCTATTTTGAAGTTTGGTGTGAGGAA 1980
QY 3144 CTCTTTAAGAACCTGAAGGAGAAACAGGAGCCCGCAGAAATGGATATAGATGAAATC 3203
Db 1981 CTCTTTAAGAACCTGAAGGAGAAACAGGAGCCCGCAGAAATGGATATAGATGAAATC 2040
QY 3204 TTCAAGAGAGCTGAAACTCGGAAATGAGCCAGGTCCTATGACTAGGGGATGAGTTG 3263
Db 2041 CTGAAGAGGCTGAAACTCGAAGAAATGAGTCAGGCTTATTAATCTAGGAGATGAGTTA 2100
QY 3264 CTCTCAGCTCAAGTGGCGAACTTTTCAATATGGATGAAGATGATATGAGTTGGAA 3323
Db 2101 CTCTCAGCTCAAGTGGCGAACTTTTCAATATGGATGAAGATGATATGAGTTGGAA 2160
QY 3324 CCAGAAGAAATTCAGAATTTGGGAAGAAATCATCCAGAAATCCCAAGGAGAGGATA 3383
Db 2161 CCAGAAGAAATTCAGAAGAACTGGGAAGAAATCATTCAGAAAGTTCAAGTGGCCAGGAATA 2220
QY 3384 GAGGAGGAGAAAGACAAAGAACTTGAAGAAATATACATGCTCCCGAGGATGAGAAAC 3443
Db 2221 GAAGAGGAGAAAGACAAAGAACTTGAAGAAATATATGCTTCCAAGATGAGAAAC 2280
QY 3444 TGTCGAAAACAGATCAGCTTTTATGGAGTGAAGGAGACGAGTAGTAGGACAGAAGATAT 3503
Db 2281 TGTCGAAAACAGATCAGCTTTTATGGAAATGAGGAGATGAGTAGGACAGAAGATAT 2340
QY 3504 TCTGGATCTGATAGTACATCCATCAGAGAAAGAAACGCCAAAAGCGTGGAGACCTT 3563
Db 2341 TCTGGATCTGATAGTATCCATCTCAGAGAAAGAAACGCCAAAAGCGTGGAGACCA 2400
QY 3564 CGAACCATTCCTCGAAGAAATATTAAGGATTTAGTGATGCAGAGATCAGCGGTTTATC 3623
Db 2401 CGAACCATTCCTCGAAGAAATATTAAGGATTTAGTGATGCAGAGATCAGCGGTTTATC 2460
QY 3624 AAGAGTTACAAGAAATTTGGTGGCCCTCTGGAAGGTTAGATGCTCTAGCTAGAGATGCT 3683
Db 2461 AAGAGTTACAAGAAATTTGGTGGCCCTCTGGAAGGTTAGATGCTCTAGCTAGAGATGCT 2520
QY 3684 GAACTGGTGAATAATCTGAGACAGACCTTAGACGCTTTGGGTGAACCTTGATACATAATGGA 3743
```

```
Db 2521 GAGCTAGTTGATAAATCTGAAACAGACCTTAGACGCTCTGGGAGAACTTGATACATAATGGA 2580
QY 3744 TGCATTAAAGGCTTTAAAGGACAAATTCATCTGGACACAAAGACGAGGAGTACAGTTGGG 3803
Db 2581 TGCATTAAAGGCTTTAAATGATAATGATCTTTGGTCAAGGAAAGACAGGTGGTAGATTTGGG 2640
QY 3804 AAAGTTAAAGGCGCAACGTTTCAATCTCAGGAGTGCAGGTGAATCAAACTAGTCATC 3863
Db 2641 AAAGTTAAAGGCGCAACATTCGAATAGCAGGATGCAGGTGAATGCAAGCTAGTCATT 2700
QY 3864 TCTCAGCAAGAGAGCTGGCACCACCTGCACAAATCATCTCTTCAGATCCAGAAAGG 3923
Db 2701 TCTCAGCAAGAGAGTTGGCACCATTGCATAAATCGATTCTCTTCAGATCCAGAAAGG 2760
QY 3924 AAAAGATATGTCATCCATCCATCCACACCAAGGCTGCTCACTTCGATATAGATTTGGGTTAA 3983
Db 2761 AAAAGATATGTCATCCATCCATCCACCAAAAGAGCTCATTTTGATATAGATTTGGGTTAA 2820
QY 3984 GAAGATGATTCCAATCTGTTAGTAGCATCTATGAATATGGCTATGGCAGCTGGGAAATG 4043
Db 2821 GAAGATGATTCCAATCTGTTAATAGGCATCTATGAATATGGTTATGGCAGCTGGGAAATG 2880
QY 4044 ATAAAATGGATCCAGATCTCAGCTTAACACAGAAAGATTTTACCTGATGATCCAGACAAG 4103
Db 2881 ATAAAATGGATCCTGATCTCAGTTTGACACAGAAAGATTTTACCTGATGATCCAGATAAG 2940
QY 4104 AAA 4106
Db 2941 AAA 2943

RESULT 6
AF181825 Aegolius funereus chromosome 2 chromodomain helicase DNA binding
LOCUS protein 1 (CHD1Z) mRNA, partial cds.
DEFINITION Aegolius funereus chromosome 2 chromodomain helicase DNA binding
ACCESSION AF181825
VERSION AF181825.1 GI:5917749
KEYWORDS Aegolius funereus.
SOURCE Aegolius funereus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauia; Aves; Neognathae; Strigiformes; Strigidae; Aegolius.
REFERENCE 1 (bases 1 to 2754)
AUTHORS Fridolfsson, A.K. and Ellegren, H.
TITLE Molecular evolution of the avian CHD1 genes on the Z and W sex
chromosomes
JOURNAL Genetics 155 (4), 1903-1912 (2000)
MEDLINE 20384598
REFERENCE 2 (bases 1 to 2754)
AUTHORS Fridolfsson, A.-K. and Ellegren, H.
TITLE Direct Submission
JOURNAL Submitted (30-AUG-1999) Department of Evolutionary Biology,
Evolutionary Biology Centre, Uppsala University, Norbyvagen 18D,
Uppsala S-75236, Sweden
FEATURES
source Location/Qualifiers
1..2754
/organism="Aegolius funereus"
/db_xref="taxon:103375"
/chromosome="2"
<1..>2754
/gene="CHD1Z"
<1..>2754
/gene="CHD1Z"
/codon_start=1
/product="chromodomain helicase DNA binding protein 1"
/protein_id="AAD56022.1"
/db_xref="GI:5917754"
/translation="ELTDLHKYOIVERITAHNSOKSAAGYPDYCKWQGLPYSECS
WEDGALIAKKFQCTCDEIDFSPNSQSTTFPKDKCVLKQRPFRVALKKQPSYIGGHEGLE
LDYQLQGLNLWLHWSKNCILADEMGLKTIQITISFLNLFHEHLYDYGFLVLP
LSTLTSWOREIQTWAPQMNAYVYLGDITSRNIRTHEMHLQTKRLKFNILLTVEIL
```

LXKSKFLGLNNAFTGVDBAHLKNDSSLKYTLDFKSNWRLLLTGTPLONSLKELM  
SLLHFIAPKFSWEDFEERHGREYGTASLHKELEFLLRRVKDVEKSLPAKVQ  
ILNEMSAQOKYIYKWLITRYKALSKSGKSTGFLNIMELKCCNCHYLILPPOG  
NEFTKQALQHLIRSSKLLILDLKLLIRKRNRLVLFQWRLMDILAEYLKYRQ  
FFQRLDGSIKGELRQALDHFNAEGSEDFCLLSTRAGGLINLASADVVIEDSM  
NPQNDLQAARHRIQKKQNIYRLVTKGSVEEDILERAKKMWLDHLVTOEMDTG  
KVLHTGSTSSSTPFNKEFUSAILFGAELFKPEGEEOPEQEMDIDILKRAETR  
ENEPGLTVGDELLSOFKVAFNSNMDEDDIELEPERNSRWEBIIPEVORRRIEER  
QKELEIYMLPRNRCAKOISFNGSERRSRYSQSDSDSISERKPKRGKRPRI  
PRENTKGSDAEIRFPIKSYKFGPLERLDAVADLVDKSETDLRLHSELVHNGC  
IKALKDSSGOERAGGRGLGVKGPTRISGVQVNAKLVISHEBELAPLHKISPDPE  
RRRVYFCHTAAHFDIWDGKEDDNLIGIYEYGVSWEMIKMDPDLSTOKILPDD  
PDKK\*

BASE COUNT 931 a 481 c 636 g 706 t

ORIGIN

Query Match 37.2%; Score 2458; DB 5; Length 2754;  
Best Local Similarity 93.3%; Pred. No. 0;  
Matches 2569; Conservative 0; Mismatches 185; Indels 0; Gaps 0;

QY 1353 GAGCTTACAGATGCTGCACAAACATATCAATAGTGGAAAGATATGCTCATTC 1412  
Db 1 GAGCTTACAGATGCTGCACAAACATATCAGATAGTGGAAAGATATGCTCATTC 60  
QY 1413 AATCAAAAGTCAGCAGCTGGTTATCCGGACTACTATTGCAAAATGGCAGGCTCGCTTAC 1472  
Db 61 AATCAGAAATCGCGGCTGGTTATCCAGACTACTATTGTAATGGCGGGCTCGCGTAT 120  
QY 1473 TCAGAAATGAGTGGGAAGATGGTCTCTCATTTGCCAAAAGTTTCAGCAGCATGAT 1532  
Db 121 TCAGAAATGAGTGGGAAGATGGTCTCTCATTTGCCAAAAGTTTCAGCAGCATGAT 180  
QY 1533 GAGTATTTTACGAGAAATCAATCCAGACTACTCCCTTTAAGGACTGCAAGGTTCTAAAA 1592  
Db 181 GAGTACTTACGAGAAATCAATCCAGACCCTCCCTTTAAGGACTGCAAGGTTCTAAAA 240  
QY 1593 CAGAGACCAAGATTTGTTGCACGAGAAAGCAACCATCTACATTTGAGGACATGAAAGT 1652  
Db 241 CAGAGGCCAAGGTTGTTGCACGAGAAAGCAACCCCTCTATATTGGAGGACATGAAAGT 300  
QY 1653 CTGGAGTTAAGAGATATCAGTTAAATGGAATGGAATGGCTCGCTCATTCATGTTGCAAA 1712  
Db 301 CTAGAAATTAAGAGATATCAATTAATGGAATGGAATGGCTCGCTCATTCATGTTGCAAA 360  
QY 1713 GGAATAGTTGATTTCTTTCGAGATGAAATGGCTCTGGGTAAACCAATACAAACAATTTCT 1772  
Db 361 GGAATAGTTGATTTCTTTCGAGATGAAATGGCTCTGGGTAAACCAATACAAACAATTTCT 420  
QY 1773 TTTCTGAACCTACCTGTTTCATGAACATCAACTGATGGCCCTTTTCTTCGCGGTGCCA 1832  
Db 421 TTTCTGAACCTACCTGTTTCATGAACATCAACTGATGGCCCTTTTCTTCGCGGTGCCA 480  
QY 1833 CTTTCTACCTTGACATCTTGGCAAGAGAGATTTCAACTTGGGCTCCTCAGATGAATGCT 1892  
Db 481 CTTTCTACCTTGACATCTTGGCAAGAGAGATTTCAACTTGGGCTCCTCAGATGAATGCT 540  
QY 1893 GTAGTTACTTAGGAGATATACCTAGTACGAATATGATAGGACTCATGATGATCAT 1952  
Db 541 GTAGTTACTTAGGAGATATACCTAGGAGATATGATAGGACTCATGATGATCAT 600  
QY 1953 CCACAGACTAAGGATTAAGTTTAACTACTTCTGAGGACATATGAAATTTTACTGAAG 2012  
Db 601 CTGACAGCCAAACGGTTAAATTTTAACTACTTCTGAGGACATATGAAATTTTACTGAAG 660  
QY 2013 GATAAGTCATCTCTGGTGGTCTCAATTTGGGCAATTCATPAGGAGTTGATGAGCTCATCGT 2072  
Db 661 GATAAGTCATCTCTGGTGGTCTCAATTTGGGCAATTCATPAGGAGTTGATGAGCTCATCGT 720  
QY 2073 TTAATAAATGATGACTCTCTCTGACAGGACTTTTAAATAGACTTTAAGTCCCAACCATCGA 2132  
Db 721 TTAATAAATGATGACTCTCTCTGACAGGACTTTTAAATAGACTTTAAGTCCCAACCATCGT 780

QY 2133 CTTCTGATTAAGTGGAAACCCCACTGCAGAAATTCCTCCTCAAGAGCTGGTCTTTGTTGAT 2192  
Db 781 CTTCTGATTAAGTGGAAACCCCACTGCAGAAATTCCTCCTCAAGAGCTGGTCTTTGTTGAT 840  
QY 2193 TTTATCATGATGCGCAAGAAATTTTCTCTCTGGGAAGATTTTGAAGAGGAGCATGGCAAGGA 2252  
Db 841 TTTATCATGATGCGCAAGAAATTTTCTCTCTGGGAAGATTTTGAAGAGGAGCATGGCAAGGG 900  
QY 2253 AGAGAGTATGGTTTATGCAAGTCTTTCACAAAGAGCTTGAACCATTTTCTACTAAGAGAGTT 2312  
Db 901 AGAGAGTATGGTTTATGCAAGTCTTTCACAAAGAGCTTGAACCATTTTCTACTAAGAGAGTT 960  
QY 2313 AAAAAAGATGTAGAAAGTCTTTTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2372  
Db 961 AAAAAAGATGTAGAAAGTCTTTTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020  
QY 2373 AGTGCATTCAGAGCAATATTTTACAAGTGGATTTTAAACAAGCAATTTATAAGCCCTCAGT 2432  
Db 1021 AGTGCATTCAGAGCAATATTTTACAAGTGGATTTTAAACAAGCAATTTATAAGCCCTCAGT 1080  
QY 2433 AAAGGTTTCAAAAGGCAAGTACCTCAGGCTTTTCTGAACATTTATGATGGAACCTTAAAGAGTGT 2492  
Db 1081 AAAGGTTTCAAAAGGCAAGTACCTCAGGCTTTTCTGAACATCATGATGGAACCTTCAAGAGTGT 1140  
QY 2493 TGTAAACATTTGCTACCTCATTAAAGCCAGAGATGATGATGATGATGATGATGATGATGATGATGAT 2552  
Db 1141 TGTAAACATTTGCTACCTCATTAAAGCCAGAGATGATGATGATGATGATGATGATGATGATGATGAT 1200  
QY 2553 GCTTTACAGCATTTGATAGTAGCAGGCGGGAACATTAATCCTTTCTGACAGCTACTGATTT 2612  
Db 1201 GCTTTACAGCATTTGATAGTAGCAGGCGGGAACATTAATCCTTTCTGACAGCTACTGATTT 1260  
QY 2613 CTTCTGAGAGAGCTGGCAACAGATTTCTGATTTCTCTCAGATGTTGAGGATGCTGAGAC 2672  
Db 1261 CTTCTGAGAGAGCTGGCAACAGATTTCTGATTTCTCTCAGATGTTGAGGATGCTGAGAC 1320  
QY 2673 ATCTAGCAGATATCTGAAGTATGCGCAGTTTCCCTTTCCAGAGACTTTGATGATGATGATGAT 2732  
Db 1321 ATCTAGCAGATATCTGAAGTATGCGCAGTTTCCCTTTCCAGAGACTTTGATGATGATGATGAT 1380  
QY 2733 AAAGGGAATTTAGGAGCAAGCAAGCTGATCATTTTCAATGCAAGAGATGATGATGATGATGAT 2792  
Db 1381 AAAGGGAATTTAGGAGCAAGCAAGCTGATCATTTTCAATGCAAGAGATGATGATGATGATGAT 1440  
QY 2793 TGTGTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2852  
Db 1441 TGTGTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1500  
QY 2853 GTAGTTATTTTGTGATTTCTGCTGGAATCCACAGATGATCTGCAAGGCAAGGAGAGCT 2912  
Db 1501 GTAGTTATTTTGTGATTTCTGCTGGAATCCACAGAGATCTTCAAGCAGGCTAGAGCA 1560  
QY 2913 CATAGATTTGAG 2972  
Db 1561 CATAGATTTGAG 1620  
QY 2973 GAAGAGATTTCTTGAAGAGCCNAG 3032  
Db 1621 GAAGAGATTTCTTGAAGAGCCNAG 1680  
QY 3033 AGAATGAG 3092  
Db 1681 AGAATGAG 1740  
QY 3093 CCTTTTAAAG 3152  
Db 1741 CCTTTTAAAG 1800  
QY 3153 GAACCTGAG 3212  
Db 1801 GAACCTGAG 1860  
QY 3213 GCTGAACTCGCGGGAATGAGCAGGTCCTTCTGCTGATGGGAGATGATGATGATGATGATGATGATGATGAT 3272



Db	1861	GCTGAACACTCGGAAATATGACCGAGTCCATTAACTGTAGAGATGAGTTGCTTTACACAG	1920
QY	3273	TTCAAGGTGCGGAACATTTTCCATATATGGATGAAGATGATATTCAGTTGGAAACACAGAAAGA	3332
Db	1921	TTCAAGGTGCGGAACATTTTCCATATATGGATGAAGATGATATTCAGTTGGAAACACAGAAAGA	1980
QY	3333	AATTCAGAATTTGGGAAGAAATCATCCAGAAATCCCAAGATCCCAAGAGAGATAGAGAGGAG	3392
Db	1981	AATTCAGAATTTGGGAAGAAATCATTCAGAAATCCCAAGATCCCAAGAGAGATAGAGAGGAG	2040
QY	3393	GAAGACAAAAGAACTTTGAAGAAATATATCATGCTCCCGAGGATGAGAACTGTGCAAAA	3452
Db	2041	GAAGACAAAAGAACTTTGAAGAAATATATCATGCTCCCAAGGATGAGAACTGTGCAAAA	2100
QY	3453	CAGATCAGCTTTAATGGGAGTGAAGGAAGACGAGTAGGAGCAGAGATATTCGGATCT	3512
Db	2101	CAGATCAGCTTTAATGGGAGTGAAGGAGGCGCAGTAGGAGCAGAGATATTCGGATCT	2160
QY	3513	GATAGTGACTCCATCAGAGAAAGAAACGCCCAAAAAGCGTGGAAAGACCTCGAACCAATT	3572
Db	2161	GATAGTGACTCCATCAGAGAAAGAAACGCCCAAAAAGCGTGGAAAGACCTCGAACCAATT	2220
QY	3573	CCTCGAGAAATATTAAGGATTTAGTGATGAGAGATCAGGCGGTTTATCAAGAGTTAC	3632
Db	2221	CCTCGAGAAATATTAAGGATTTAGTGATGAGAGATCAGGCGGTTTATCAAGAGTTAC	2280
QY	3633	AAGAAATTTGGTGGCCCTCTGGAAGGTTAGATGCTGTAGCTAGATGAGATGCTGAACGTT	3692
Db	2281	AAGAAGTTTGGTGGCCCTCTGGAAGGTTAGATGCTGTAGCTAGATGAGATGCTGAACGTT	2340
QY	3693	GATAAATCTCGAGACAGACCTTAGACGCTTTGGGTGAACCTGTACATATGATGATGCATTAAG	3752
Db	2341	GATAAATCAGACAGACAGACCTTAGACGCTTTGGGAGAGCTTGTACATATGATGATGCATTAAG	2400
QY	3753	GCTTTAAGAGCAATTCATCTGCAGAAAGAAAGCAGGAGGTAGACTTGGGAAGTTAAA	3812
Db	2401	GCTTTAAGAGCAATTCATCTGCAGAAAGAAAGCAGGAGGTAGACTTGGGAAGTTAAA	2460
QY	3813	GGCCCAACGTTTCGAATCTCAGGAGTGACGTTGAATGCAAAACTAGTTCATCTCTCAGGAA	3872
Db	2461	GGCCCAACATTCGATCTCTGAGTGACGTTGAATGCAAAACTAGTTCATCTCTCAGGAA	2520
QY	3873	GAAGAGCTGGCACCCTGCACAAATCCATTCCTCAGATCCAGAGAAAGGAAAGATAT	3932
Db	2521	GAAGAGTTGGCACCCTGCACAAATCCATTCATCCATCAGATCCAGAGAAAGGAAAGATAT	2580
QY	3933	GTATCCCATGGCCACACCAAGGCTGCTCAGTTCGATATAGATTGGGTAAGAGATGAT	3992
Db	2581	GTATCCCATGGCCACACCAAGGCTGCTCAGTTCGATATAGATTGGGTAAGAGATGAT	2640
QY	3993	TCCAATCTGTTAGTATGATATGAATATGGCTATGGCAGCTGGGAATGATATAAATG	4052
Db	2641	TCCAATCTGTTATAGGATCTATGATATGGTTAGGTCGGCAGCTGGGAATGATATAAATG	2700
QY	4053	GATCCAGATCTCAGCTTAACACAGAGATTTTACCTGATGATCCAGACAGAAA	4106
Db	2701	GATCCCTGATCTCAGTTTGACACAGAGATCTTGCCTGATGATCCAGATAAGAAA	2754
RESULT	7		
AF181828			
LOCUS		2754 bp	mRNA linear VRT 07-AUG-2001
DEFINITION			Nymphicus hollandicus chromosome 2 chromodomain helicase DNA binding protein 1 (CHD12) mRNA, partial cds.
ACCESSION	AF181828		
VERSION	AF181828.1	GI:5917752	
KEYWORDS			
SOURCE			cockatiel.
ORGANISM			Nymphicus hollandicus
			Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Psittaciformes; Cacatuidae; Nymphicus.

REFERENCE	1 (bases 1 to 2754)		
AUTHORS	Fridolfsson,A.K. and Ellegren,H.		
TITLE	Molecular evolution of the avian CHD1 genes on the Z and W sex chromosomes		
JOURNAL	Genetics 155 (4), 1903-1912 (2000)		
MEDLINE	20384598		
REFERENCE	2 (bases 1 to 2754)		
AUTHORS	Fridolfsson,A.-K. and Ellegren,H.		
TITLE	Direct Submission		
JOURNAL	Submitted (30-AUG-1999) Department of Evolutionary Biology, Evolutionary Biology Centre, Uppsala University, Norbyvagen 18D, Uppsala S-75236, Sweden		
FEATURES	Location/Qualifiers		
source	1..2754		
	/organism="Nymphicus hollandicus"		
	/strain="Lutino cockatiel"		
	/db_xref="taxon:13180"		
gene	/chromosome="Z"		
	<1..>2754		
CDS	/gene="CHD12"		
	<1..>2754		
	/gene="CHD12"		
	/codon_start=1		
	/product="chromodomain helicase DNA binding protein 1"		
	/protein_id="AAD56025.1"		
	/db_xref="GI:5917752"		
	/translation="ELTDLHKQYQIIVERIIAHSNOKSAAGYDHYCKWQGLPYSFCS WEDGALIAKFOARIDEYFSRNQSKTPFKDCKVLKQRPFRVALKKQPLYIGGHESLE LKDYQLNGLNWLHWSKNCILADEMGIGKTQITISFNILFHEQLYGPFLVVP LSTLSWREIQITWASQMNNAVYLDITSRNMIRTHEWMHPQTKRLKFNILLTYEIL LKDKSFLGGLNMAFIVGDAHRLLKNDLSLLYKTLIDFKSNHRLITGTPLONSKEILW SLUHTFMPEKFSFWEGBFEHKGREYGVASLHKELEPFLLRVKDKVKSIPAKVEQ ILRMSALQKOYKWLTRYKALKSGSKSGTSGFLNIMMELKCCNHYLKKPPDD NEFYAKQEAQLHLIRSSGKLILLKLLRLRGNRVLIFSQVMRMLDILAEVLKYRQ FPFRLDGSIGELRKQALDHFNAEGSEDFCLLSTRAGGLGINLASADIVLFDSDM NPQNDLQAQRAHRIQKQVNIYRLVTGKSVVEEDILERAKKMVLHDHVLQRMDDTG KTVLHTGSTPSSPTFNKEELSAILKFGAEELFKEPEGEQEPQEMDIDELIKRAETR ENEPGLTYVDELLSQFKVAFNSNMDDEIDLEPNCRNWBEIIPVORRRKEEER OKELEIYVLMRMCNAKOISFNGSEGRSRYSRSGSDSDISERKRKKRPRPTI PRENTKFSDAETIRRFIKYKFGGPLERLDAVARDAELVDKSEIDLRLRIGELVHNGC IKALDSSSGQERAGRLGKVGKPTFRISGVQVNAKLVISHEEELAPLHKSIPSPDEE RRYVTPCHTKAAHFDIDMGKEDDNLNLIIGIYEYGYGSWEMKMDPDLSLUTOKILPDD PDK"		
BASE COUNT	926 a 470 c 642 g 716 t		
ORIGIN			
Query Match	37.2%;	Score 2458;	DB 5; Length 2754;
Best Local Similarity	93.3%;	Pred. No. 0;	
Matches 2569;	Conservative 0;	Mismatches 185;	Indels 0; Gaps 0;
QY	1353	GAGCTTTACAGATCATCTGCACAAACAATATCAAAATAGTGGAAAGAATAATTGCTCATTTCA	1412
Db	1	GAGCTTTACAGACACTCTACATAACAACATATCAGATAGTGGAAAGAATAATTGCTCATTTCA	60
QY	1413	AATCAAAAGTACAGAGCTGGTTATCCGGACTACTATTGCAAAATGGCAGGCTCTGCCTTAC	1472
Db	61	AATCAGAAATCAGCAGCTGGTTATCCAGACCACTATTGTTAAATGGCAGGCTCTGCCCTAT	120
QY	1473	TCAGAATGTAGCTGGGAAGATGCTGCTCATTTGCCAAAAAGTTTCAGGCACCATTTGAT	1532
Db	121	TCAGAATGTAGCTGGGAAGATGCTGCTCATTTGCCAAAAAGTTTCAGGCACCATTTGAT	180
QY	1533	GAGTATTTTATGACAGAATAATCAATCAAGACTACTCCCTTTTAAAGGACTGCAAGGTTCTAAAA	1592
Db	181	GAGTACTTTTATGACAGAATAATCAATCAAGACTACTCCCTTTTAAAGGACTGCAAGGTTCTAAAA	240
QY	1593	CAGAGACCAAGATTTGTTGCACCTGAAGAACCAACACTCTTACATTTGAGGACATGAAAGT	1652
Db	241	CAGAGCCCAAGGTTTCTTGCACCTGAAGAACCAACACTTGTACATTTGGGGACATGAAAGT	300
QY	1653	CTGGAGTTAAGAGATTTACGTTAAATGATTGAATGGCTCGCTCATTCATTCGTCGCAAA	1712



Db 301 TTGGAGTTAAGAGATTATCAGTTAAATGGATTGAATTTGGCTTGGCTCACTCATCTGGTGCAAA 360  
Qy 1713 GGAAATAGTTGTTATCTTTCAGATGAAATGGGTCTGGGTAAACAAATACAAACAAATTTCT 1772  
Db 361 GCGAATAGCTGTATCTTTCGCGATGAAATGGGTCTGGGTAAACAAATACAAACAAATTTCT 420  
Qy 1773 TTCTCGAATACCTGTTTCATGAACATCAACTGTATGSCCTTTTCTTCTCGCGTGCCTCA 1832  
Db 421 TTCTCGAATACCTGTTTCATGAACATCAATTTGATGGCTTTCTTCTGCTGCTGCTGCCA 480  
Qy 1833 CTTTCTACCTTGCATCTTGGCAAGAGAGATTCAAACTTGGCTTGGCTCCTCAGATGAATGCT 1892  
Db 481 CTTTCTACCTTGCATCTTGGCAAGAGAGATTCAAACTTGGCTTGGCTCCTCAGATGAATGCC 540  
Qy 1893 GTAGTTTACTTAGGAGATATACTAGTATGAATATGATAAGGACTCATGAATGGATGCAT 1952  
Db 541 GTAGTTTACTTAGGAGATATACTAGCAGAAATATGATAGAACTCATGAATGGATGCAT 600  
Qy 1953 CCACAGACTAAACGATTAAAGTTTAACTACTTCTGACGACATATGAAATTTTACTGAAG 2012  
Db 601 CCACAGACCAACGATTAAATTTTAACTACTTCTTAACAGCTATGAATCTTCTTTAG 660  
Qy 2013 GATAGTCATCTCTGGTGGTCTCAATTTGGGCAATTCATAGAGTTGATGAAGCTCATCGT 2072  
Db 661 GATAGTCATCTCTGGTGGTCTCAATTTGGGCGTTTATAGAGTTGATGACGAGCTCATCGT 720  
Qy 2073 TTAATAATGATGACTCTCTCTGACAGACTTTTAAATAGACTTTAAGTCCAAACCATCGA 2132  
Db 721 TTGAAATGACGACTCTCTCTGACAGACTTTTAAATAGACTTTAAGTCTAAACCATCGT 780  
Qy 2133 CTTCTGATTACTGGAAACCCACTGCAAAATTCCTCAAGAGCTGTGGTCTTTGTGTCAT 2192  
Db 781 CTTCTGATTACTGGAAACCCCTTTGCAAACTCCCTGAAAGAGCTTTGGTCTTTGTGTAC 840  
Qy 2193 TTCATCATGCCAGAAAATTTTCTCTCTGGGAAGATTTTGAAGAGGAGCATGGCAAAAGGA 2252  
Db 841 TTCATCATGCCAGAAAAGTTTCTCTCTGGGAAGTTTGAAGAGGAGCATGGCAAAAGG 900  
Qy 2253 AGAGAGTATGGTTATGCAAGTCTTTCACAAAGAGCTTGAACCATTTTACTAAGAGAGTT 2312  
Db 901 AGAGAGTATGGTTATGCAAGCTTTCATFAAGAACTTGAACCATTTCTGCTGAGAGAGTT 960  
Qy 2313 AAAAAGATGAGAAAAGTCTTTACCTGCTAAAGTTGAACAAAATCTGAGGATGAAATG 2372  
Db 961 AAAAAGATGAGAAAAGTCTTTACCTGCTAAAGTTGAGCAAAATCTGAGGATGAAATG 1020  
Qy 2373 AGTCGATTGCAAGCAATATTACAAGTGGATTTTAAACAGGAATTAACAGCCCTCAGT 2432  
Db 1021 AGTCGATTGCAAGCAATATTACAAGTGGATTTTAAACAGGAATTAACAGCCCTCAGT 1080  
Qy 2433 AAAGTTCAAAAGCAGTACCTCAGGCTTCTGAACATTATGATGAAGTTAAGAGTGT 2492  
Db 1081 AAAGTTCAAAAGCAGTACCTCAGGCTTTTGAACATCATGATGAGACTCAAGAGTGT 1140  
Qy 2493 TGTAAACATTTGCTTACCTCATTTAAGCCACAGATGATGAATGAATCTATTAACACAGAG 2552  
Db 1141 TGTAAACATTTGCTTACCTCATTTAAACACAGATGATGAATGAATTTTATTAACACAGAG 1200  
Qy 2553 GCCTTACAGCATTTGATACGATGAGCAGCGGGAACATAATCCTTCTTGAACAGCTACTGATT 2612  
Db 1201 GCCTTACAGCATTTAATACGATGAGCAGTGGGAACATAATCCTTCTTGAACAGCTACTGATT 1260  
Qy 2613 CGTCTGCGAGAGCTGGCAACAGAGTTCTGATTTTCTCTCAGATGGTGAAGTCTGGAC 2672  
Db 1261 CGTCTGCGAGAGCTGGCAACAGAGTTCTAATTTTCTCAGAGATGGTGAAGTCTGGAC 1320  
Qy 2673 ATCTTAGCAAAATATCTGAAGTATGCGCAGTTTCCCTTCCAGAGACTTGTGATGATCAATA 2732  
Db 1321 ATCTTAGCAAAATATCTGAAGTATGCTGATCTGATTTCCCTTTCAAGACTTGAAGTCAATA 1380  
Qy 2733 AAAGGGAATTTGAGGAAGCAAGCAGTGGATCATTTTCAATGCAAGAGATCAGAGGATTC 2792  
Db 1381 AAAGGGAATTTGAGGAAGCAAGCAGTGGATCATTTTATGCAAGAGTTTCAGAGGATTC 1440

Qy 2793 TGTTTTTTACTGTCTACAAAGAGCTGGAGGATTAGGTATTAACTTTGGCATCTCTGCACACT 2852  
Db 1441 TGTTTTTTACTGTCTACCAAGAGCTGGAGGATTAGGTATTAACTTTGGCATCTCTGCACACT 1500  
Qy 2853 GTAGTTATTTTGAATTTCTGACTGGAATCCACAGAATGATCTGCAGGCACAGCGAGAGCT 2912  
Db 1501 GTAGTTATTTTGAATTTCTGACTGGAATCCACAGAATGATCTGCAGGCACAGCGTAGACA 1560  
Qy 2913 CATAGAATTGGCAGAGAAACAGGTTAATTTATTCGCTAGTCAAAAGGATCAGTA 2972  
Db 1561 CATAGAATTGGCAGAGAAACAGGTTAATTTATTCGCTAGTCAAAAGGATCAGTA 1620  
Qy 2973 GAAGAATATTTCTTGAAGAGCCAAAGAGAGTGGCTGAGACATTTAGTAATTCAG 3032  
Db 1621 GAAGAATATTTCTTGAAGAGCCAAAGAGAGTGGTGTAGATCATCTAGTAATTCAG 1680  
Qy 3033 AGAATGGACAGACAGAAAACCTTCTGCATACAGGTTCACTCATCAAGCTTACA 3092  
Db 1681 AGAATGGACACTACAGGAAAACCTGCTACTACTACAGGCTCTACTCTTCAAGCTCAACA 1740  
Qy 3093 CTTTTTAATAAGAGAGTTTATCAGCTTATTTTGAAGTTTGGTGGCTGAGAACTCTTTAAA 3152  
Db 1741 CTTTTTAATAAGAGAGTTTATCCCGGATTTTGAATTTTGGTGGCTGAGAACTTTTTAAA 1800  
Qy 3153 GAACCTGAAGGAGAAACAGAGCCCGAGGAAATGGATATAGATGAATCTTTGAAGAGA 3212  
Db 1801 GAACCTGAAGGAGAAACAGAGCCCGAGGAAATGGATATAGATGAATCTTTGAAGAGG 1860  
Qy 3213 GCTGAACCTCGGGAAAATGAGCCAGGTCCATTTGACTGTAGGGGATGAGTTGCTTTTCACAG 3272  
Db 1861 GCGGAACCTCGGGAAAATGAGCCAGGTCCATTTAACTGTGGGAGATGAATTTGCTTTTCACAG 1920  
Qy 3273 TTCAAGGTTGCGCAACTTTTCCAAATATGGATGAAGATGATATTGAGTTTGGAAACACAGAAA 3332  
Db 1921 TTCAAGGTTGCGCAACTTTTCCAAATATGAGCAAGATGATATTGAGTTTGGAAACACAGAA 1980  
Qy 3333 AATTCAGAAATTTGGGAAGAAATCATCCAGAAATCCCAACGAGAGATGAGAGGAGG 3392  
Db 1981 AATTCAGAAATTTGGGAAGAAATCATTTCCAGAAATCCAGCGCGCAAGATAGAGAGGA 2040  
Qy 3393 GAAAGACAAAAGAACTTTGAAGAAATATACATCTCCGAGAGATGACAAACTTGTCAAAA 3452  
Db 2041 GAAAGACAAAAGAACTTTGAAGAAATCTACATCTCCGAGATGACAAACTTGTCAAAA 2100  
Qy 3453 CAGATCAGCTTTTATGGGAGTCAAGAGAGCGCATGAGGACGAGATATTTCTGATCT 3512  
Db 2101 CAGATCAGCTTTTATGGAGTCAAGGAGCGGAGTGAAGTAGAAGATCTCTGATCT 2160  
Qy 3513 GATAGTACTCCATCAACAAAAGAAAGCGCCAAAAGCGTGAAGACTCGAAGCAAT 3572  
Db 2161 GATAGTACTCCATCTCAAAAAGGAAAGCGCCAAAAGCGTGAAGCAAGCAAGCAAT 2220  
Qy 3573 CTTCCAGAAAATATTAAGAGATTTAGTGTGAGAGATCAGCGGTTTATCAAGAGTTAC 3632  
Db 2221 CTTCCAGAAAATATTAAGAGATTTAGTGTGAGAGATCAGCGGTTTATCAAGAGTTAC 2280  
Qy 3633 AAGAAATTTGGTGGCCCTCTGGAAGGTTAGTGTGTAGTAGAGATGCTGAAGTGT 3692  
Db 2281 AAGAAATTTGGTGGCCCTCTTGAAGGTTAGTGTGTAGTAGAGATGCTGAAGTGT 2340  
Qy 3693 GATAATCTGAGACAGACCTTAGAGCTTTGGGTGAACCTTGTACATAATGGATGCATTAAG 3752  
Db 2341 GATAATCTGAGACAGATAGATCTTAGAGCTTTGGGAGAGCTTGTACATAATGGATGCATTAAG 2400  
Qy 3753 GCTTTAAAGAGCAATTCATCTGGCAAGAAAGAGCAGGAGGTAGATCTTGGGAAAGTTAAA 3812  
Db 2401 GCTTTAAAGAGCAATTCATCTGGCAAGAAAGAGCAGGAGGTAGATCTTGGGAAAGTTAAA 2460  
Qy 3813 GSCCAAGCTTTGCAATCTCAGAGTGCAGTCAATGCAAACTAGTCTCTCTCACGAA 3872  
Db 2461 GSCCAAGCTTTGCAATCTCAGAGGTCAGGCTGAAGTGAAGTGAAGTCTCTCTCATGAA 2520





```
/strain="Lutino cockatiel"
/db_xref="taxon:13180"
/chromosomes="w"
<!.>2754
/gene="CHD1w"
CDS
<!.>2754
/gene="CHD1w"
/codon_start=1
/product="chromodomain helicase DNA binding protein 1"
/db_xref="GI:5917756"
/translation="ELTDLHKQYQIVERTIAHSNQSSEAGYPDYKQWGLPYSECS
WEDGALIAKKQACIDEXFSRNLSTPFDKILRQPRFVALKQPSYIGGHESLE
LRDYQLGLNLWLHSGKNSCILLADEMGLKTIQTIISFMYLFEHQLYGPFLVVP
LSTLSWOREIQITWAPOMNAVYVLDITSRNIRTHWHPOTRKLKNILLTVEIL
LKDKSFLGLNWEIGVDEAHLKNDLSLYKTLIDFSNHRLLITGTPLONSLRELW
SLLHKIMEKSSWEDFEERHKGREFGYASHLKELEPFLLRVKKDVEKSLPAKVEQ
ILRMSALQOHLIRSSGLKLLLDKLLIRERGNRVLIFSOMVRLDILAEYLKRYQ
NFEYNKQALQHLIRSSGLKLLLDKLLIRERGNRVLIFSOMVRLDILAEYLKRYQ
FPQRLDGS IKGELRKQALDHFNAGSEDFCLLSTRAGGLGINLASDVTYVFDSDW
NPQDLQQAARHRTGQKQVNIYELVTGKSEVEDILERAKKMKVLDHLVQIORMDTTG
KTVLHTGSTPSSSTPFNKEELSAIILFGAELEFKPEGEERPEQEMDIDELIKRAETR
ENEPGITVGEELLQFKVANSFNWDEDDILEPQNLNWEELIPEVORRRIEER
QKELKIIWLPWRMCAKQISFNGSRHRSRRYSGSDSDSISERRPRKRGPRPTI
PRENTKGFSDAIEIRFIKSYKFGGFLERLDAIARDLVKSDTDRLRGLGVHNGC
IKALNDNFSVQERAAAGRLGVKGPTRISGVQVNAKLVIISHEEELAPLHKSPSPDEE
RRKRYVPCHTRAAHPDIDWCKEDDSNLLIGIYEYGYGSEWEMIKMDPDLSTQKILPDD
PDKK"
BASE COUNT 947 a 464 c 613 g 730 t
ORIGIN
Query Match 36.4%; Score 2406.8; DB 5; Length 2754;
Best Local Similarity 92.1%; Pred. No. 0;
Matches 2537; Conservative 0; Mismatches 217; Indels 0; Gaps 0;
QY 1353 GAGCTTACAGATGATCTGCACAAACAATATCAAAATAGTGGAAAGATAATATGCTCATTTCA 1412
Db 1 GAACCTTACAGATGATCTACATAAACAATATCAAAATAGTGGAAAGATAATATGCTCATTTCA 60
QY 1413 AATCAAAAGTCAGACGCTGGTATCCGGACTACTATTGCAAAATGGCAGGCTGCGCTTAC 1472
Db 61 AATCAGAAATCAGAACCTGGTATCCAGATTAATGCAAAATGGCAGGCTGCGCTTAC 120
QY 1473 TCAGAAATGATGGGAAGATGGTGTCTCTCATTTGCCAAAAGATTTTCAGGACGCGATTTGAT 1532
Db 121 TCAGAAATGATGGGAAGATGGTGTCTCTCATTTGCCAAAAGATTTTCAGGACGCTGATTTGAT 180
QY 1533 GAGTATTTAGCAGAAATCAATCCAGACTACTCCCTTTAAGGACTGCAAGGTTCATAAA 1592
Db 181 GAGTATTTAGCAGAAATCTATCCAAGACCACTCCATTTAAGGACTGCAAGATTTCTAAGA 240
QY 1593 CAGAGACCAAGATTTCTGCACCTGAAGAAGACCACTATTCATTTGGAGGACATGAAGT 1652
Db 241 CAGAGCCGAGGTTTGTTGCACCTAAGAAGAACCACTATTTATTGGAGGCAATGAGAGT 300
QY 1653 CTGGAGTTAAGGATTTACAGTTAAATGGATTGAATTTGGCTCGCTCATTCATCGGTGCAAA 1712
Db 301 TTAGAGTTAAGGATTTACAGTTAAATGGACTGAACCTGCTGCTCATTCATCGGTGCAAA 360
QY 1713 GGAATAGTTGTTATCTTCAGATGAATAGGTTGGGTCTGGGTAAACAATACAAATTTCT 1772
Db 361 GGAATAGTTGTTATCTTCAGATGAATAGGTTGGGTCTGGGTAAACAATACAAATTTCT 420
QY 1773 TTCTGAACCTACTGTTTCATGAACATCAACTGATAGGCCCTTTTCTCTGCGCGTGCCA 1832
Db 421 TTATGAACCTACTGTTTCACGAACATCAACTGATAGGCCCTTTTCTGCTAGTCGTGSCCA 480
QY 1833 CTTTCTACCTTGACATCTTTGGCAAGAGAGATTCAAATTTGGGCTCCTCAGATGAATGCT 1892
Db 481 CTTTCTACCTTGACATCTTTGGCAAGAGAGATTCAAATTTGGGCTCCTCAGATGAATGCT 540
QY 1893 GTAGTTTACTTAGGAGATATAACTAGTAGAAATATGATAAGGACTCATGAATGGATGCAAT 1952
```

```
Db 541 GTTGTTTACTTAGGAGATATAACTAGCAGAAATATGATAAGAACTCATGAATGGATGCAT 600
QY 1953 CCACAGACTAAAGGATTAAGATTTAAACATACCTTCTCAGCAGCATATGAATTTTACTGAAG 2012
Db 601 CCACAGACAAAAGGATTAAGATTTAAATATACCTTCTTAACAACTTATGAATTTTGTGCAAG 660
QY 2013 GATAAGTCATTCCTTGGTGGTCTCAATTTGGCATTTCATAGGAGTTGATGAAGCTCATCGT 2072
Db 661 GATAAGTCATTCCTTGGTGGTCTGAATTTGGGTATTTATAGGAGTTGACGAAGCTCATCGT 720
QY 2073 TTAATAAATGATAGCTCTCTTCTGTACAGACTTTTAATAGACTTTAAGTCCAAACCATCGA 2132
Db 721 TTAATAAATGATAGCTCTCTTCTGTATTAAGACTTTTAATAGACTTTAAGTCTAATCATCGT 780
QY 2133 CTTCTGATTACTTGGAAACCCACTGCAAAATTCCTCTCAAAAGCTGTGCTTTTGTGGCAT 2192
Db 781 CTTCTTATTACTTGGAAACCCCTCTACAAAATTCCTCTCAGAGAGCTTTGGTCTTTTGTGTCAC 840
QY 2193 TTCATCATGCCAGAAAATTTTCTCTCTCGGGAAGATTTTGAAGAGGAGCATGGCAAGGA 2252
Db 841 TTCATCATGCCAGAAAATTTTCTCTCTCGGGAAGATTTTGAAGAGGAGCATGGCAAGGG 900
QY 2253 AGAGATATGTTATGCAAGCTCTTCAAAAGAGCTTTGAACCATTTTACTAAGAAGATT 2312
Db 901 AGAGAGTTTGGTTATGCAAGCTCTTCAAAAGAACTTTGAACCATTTTACTCAGAAGATT 960
QY 2313 AAAAAAGATGTAGAAAGTCTTTTACCTGCTAAGTGTGAACAAATCTCGAGGATGGAAATG 2372
Db 961 AAAAAAGATGTAGAAATATCTTTTACCTGCCAAGTTTGAACAGATTTTGAGGATGGAAATG 1020
QY 2373 AGTGATTCAGAGAACAATATTACAAAGTGGATTTTAAACAGGAATTTAAAGCCCTCAGT 2432
Db 1021 AGTGATTCAGAGAACAATATTACAAAGTGGATTTTAAACAGGAATTTAAAGCCCTGAGT 1080
QY 2433 AAGGTTTCAAAAGGAGTACCTCAGGCTTTCTGAAACATTTATGATGGAACCTTAAGAAGTGT 2492
Db 1081 AAGGTTTCAAAAGGAGTACCTCAGGATTTCTGAACATCATGATGGAACCTCAAGAAGTGT 1140
QY 2493 TGTAAACATTTGCTTACCTCATTAAGCCACAGATGATTAAGTATCTTATATAAACAGGAG 2552
Db 1141 TGTAAACATTTGCTTACCTCATTAAGCCACAGATGATTAAGTATCTTATATAAACAGGAG 1200
QY 2553 GCCTTACAGCATTTGATACGTAGCAGCGGAACTAATCCTTCTTGACAAGCTACTGATT 2612
Db 1201 GCCTTACAGCATTTAATACGTAGCAGTGGGAACTAATCCTTCTTGACAAGCTACTGATT 1260
QY 2613 CGTCTCGGAGAGCTGGCAACAGAGTTCTGATTTTCTCTCAGATGGTGAGGATGCTGGAC 2672
Db 1261 CGTCTACGAGAAGCTGGCAACAGAGTACTGATTTTCTCTCAGATGGTGAGAAATGCTAGAC 1320
QY 2673 ATCTTAGCAGAATATCTGAAGTATCCAGTATTCCTCTCCAGAGACTTGATGATCAATA 2732
Db 1321 ATCTTAGCAGAGATTTTGAAGTATCTGATGATTCCTCTCCAGAGACTTGATGATCAATA 1380
QY 2733 AAAGGGGAATGAGGAAAGCAAGCACTGGATCAATTTCAATGCAGAAGGATCAGAGGATTC 2792
Db 1381 AAAGGGAATGAGGAAAGCAAGCCCTGGATCAATTTTAAATGCAGAAGGATCAGAGGACTTC 1440
QY 2793 TGTTTTACTGCTACAGAGCTGGAGGATTAGGTATTAATTTGGCATCTGCTGACACT 2852
Db 1441 TGTCTTCTGCTACAGAGCTGGAGGATTAGGTATTAATTTGGCATCTGCTGACACT 1500
QY 2853 GTAGTTTATTTTGTATCTGATCGAATCCACAGAATGATCTGCAGGCACAGGCGAGAGCT 2912
Db 1501 GTAGTTTATTTGTATCTGATCGAATCCACAGAATGATCTGCAGGCACAGGCTAGAGCT 1560
QY 2913 CATAGAATTTGGACAGAAAGAAACAGGTTAATATTTATTCGGCTAGTCAACAAAGGATCAGTA 2972
Db 1561 CATAGGATTTGGACAAAGAAACAGGTTAATATTTATTCGGCTAGTCAACAAAGGATCAGTA 1620
QY 2973 GAAGAAGATATTCTTGAAGAGCCCAAGAGATGGTGTCTAGACCATTTAGTAATTCAG 3032
```



Db 125 CTA - GTTGAACAAATTTCTGAGGATGGAAATGATGCTATTCGACAGCAATATTACAAGT 183  
QY 2401 GGATTTTAAACAAGGAATTTATAAGCCCTCAGTAAAGGTTTCAAAAGGCGAGTACCTCAGGCT 2460  
Db 184 GGATTTTAAACAAGGAATTTATAAGCCCTCAGTAAAGGTTTCAAAAGGCGAGTACCTCAGG - 242  
QY 2461 TTCTGAACATTTATGATGGAATCTTAAGAAGTGTTGTAACCAATTCCTACCTCATTTAAGCCAC 2520  
Db 243 TTCTGAACATTTATGATGGAATCTTAAGAAGTGTTGTAACCAATTCCTACCTCATTTAAGCCAC 302  
QY 2521 CAGATGATATGAATTTCTATATAATAACACAGAGCCCTTACAGCAATTTGATACGTAGCAGCG 2580  
Db 303 CAGATGATATGAATTTCTATATAATAACACAGAGCCCTTACAGCAATTTGATACGTAGCAGCG 362  
QY 2581 GGAACATAATCTCTTGAACAGCTACTGATTCGCTCGCAGAACGTTGGCAACACAGAGTTC 2640  
Db 363 GGAACATAATCTCTTGAACAGCTACTGATTCGCTCGCAGAACGTTGGCAACACAGAGTTC 422  
QY 2641 TGATTTTCTCTCAGATGGTGAGGATGCTGGACATCTCTAGCAGAAATATCTGAAATATCGCC 2700  
Db 423 TGATTTTCTCTCAGATGGTGAGGATGCTGGACATCTCTAGCAGAAATATCTGAAATATCGCC 482  
QY 2701 AGTTTCCCTTCCAGAGACTTGATGGATCAATAAAGGGGAATTTGAGAACCAAGCACTGG 2760  
Db 483 AGTTTCCCTTCCAGAGACTTGATGGATCAATAAAGGGGAATTTGAGAACCAAGCACTGG 542  
QY 2761 ATCATTTCAATCGAGAGACTCAGAGGATTTCTGTTTTTTTACTGCTCTACAGAGCTGGAG 2820  
Db 543 ATCATTTCAATCGAGAGACTCAGAGGATTTCTGTTTTTTTACTGCTCTACAGAGCTGGAG 602  
QY 2821 GATTAGGATTAATCTTGGCAATCTGCTGACACTGTAGTTATTTTGTATTCGTACTGGAATC 2880  
Db 603 GATTAGGATTAATCTTGGCAATCTGCTGACACTGTAGTTATTTTGTATTCGTACTGGAATC 662  
QY 2881 CACAGATGATCTCGAGGCACGCGAGACTCATAGAAATTTGGACAGAAACAGGTTA 2940  
Db 663 CACAGATGATCTCGAGGCACGCGAGACTCATAGAAATTTGGACAGAAACAGGTTA 722  
QY 2941 ATATTTATCGGCTAGTCACAAAAGGATCAGTAGAAGAAGATATTTCTTGAAGAGCCAAAGA 3000  
Db 723 ATATTTATCGGCTAGTCACAAAAGGATCAGTAGAAGAAGATATTTCTTGAAGAGCCAAAGA 782  
QY 3001 AGAAGATGGTGCTAGACCAATTTAGTAATTCAGAGAATGGACACAGGAAAAAATCTTTC 3060  
Db 783 AGAAGATGGTGCTAGACCAATTTAGTAATTCAGAGAATGGACACAGGAAAAAATCTTTC 842  
QY 3061 TGCATACAGCTTCAACTCCATCAAGCTCTACACTTTTATAAGAGAGTATTACAGCTA 3120  
Db 843 TGCATACAGCTTCAACTCCATCAAGCTCTACACTTTTATAAGAGAGTATTACAGCTA 902  
QY 3121 TTTTGAAGTTTGGTGTGAGGAATCTTTTAAAGAACTGAAGGAGAGAACAGAGAGCCCC 3180  
Db 903 TTTTGAAGTTTGGTGTGAGGAATCTTTTAAAGAACTGAAGGAGAGAACAGAGAGCCCC 962  
QY 3181 AGGAATGGATATAGATGAAATCTTGAAGAGAGCTGAAACTCGGGAAAAATPGACCCAGGTC 3240  
Db 963 AGGAATGGATATAGATGAAATCTTGAAGAGAGCTGAAACTCGGGAAAAATPGACCCAGGTC 1022  
QY 3241 CATTTGAGTGTAGGGATGAGTTCGTTTACAGTTCAAGGTGCGCACTTTTCCAATATGG 3300  
Db 1023 CATTTGAGTGTAGGGATGAGTTCGTTTACAGTTCAAGGTGCGCACTTTTCCAATATGG 1082  
QY 3301 ATGAAGATGATTTAGTTTGAACCAAGAAAGAAATTCAGAAATTTGGAGAGAAATCATCC 3360  
Db 1083 ATGAAGATGATTTAGTTTGAACCAAGAAAGAAATTCAGAAATTTGGAGAGAAATCATCC 1142  
QY 3361 CAGAAATCCCAACGGAGAGGATAGAGGAGGAGAAAGACAAAAGAACTTTGAAGAAATAT 3420  
Db 1143 CAGAAATCCCAACGGAGAGGATAGAGGAGGAGGAGAAAGACAAAAGAACTTTGAAGAAATAT 1202  
QY 3421 ACATGCTCCCGAGGATGAGAAACTGTGCAAAACAGATCAGCTTTAATGGAGTGAAGGAA 3480  
Db 1203 ACATGCTCCCGAGGATGAGAAACTGTGCAAAACAGATCAGCTTTAATGGAGTGAAGGAA 1262

QY 3481 GACCCAGTAGGAGCAGAGAATATTTCTGGATCTGATAGTACTCCATCACAAGAAAC 3540  
Db 1263 GACCCAGTAGGAGCAGAGAATATTTCTGGATCTGATAGTACTCCATCACAAGAAAC 1322  
QY 3541 GGCACAAAAGCGTGAAGACCTCGAACCATTTCTCGAGAAAAATATTAAGGATTTAGTG 3600  
Db 1323 GGCACAAAAGCGTGAAGACCTCGAACCATTTCTCGAGAAAAATATTAAGGATTTAGTG 1382  
QY 3601 ATCAGAGATCAGCGGTTTATCAAGAGTTTACAAGAAATTTGGTGGCCCTCTGAAAAGGT 3660  
Db 1383 ATCAGAGATCAGCGGTTTATCAAGAGTTTACAAGAAATTTGGTGGCCCTCTGAAAAGGT 1442  
QY 3661 TAGATGCTGTAGCTAGAGATGCTGAATGTTGATTAATCTTGAGACAGACTTTAGAGGTT 3720  
Db 1443 TAGATGCTGTAGCTAGAGATGCTGAATGTTGATTAATCTTGAGACAGACTTTAGAGGTT 1502  
QY 3721 TGGGTGAATGTTGATACATAATTAAGGATTAAGGCTTTTAAAGGACAAATTCATCTGGACAAG 3780  
Db 1503 TGGGTGAGCTTGTACATAATTAAGGATTAAGGCTTTTAAAGGACAAATTCATCTGGACAAG 1562  
QY 3781 AAAGACAGAGGCTAGACTTTGGGAAAGCTTAAAGGCCCAAGCTTTTGAATCTCAGGAGTGC 3840  
Db 1563 AAAGACAGAGGCTAGACTTTGGGAAAGCTTAAAGGCCCAAGCTTTTGAATCTCAGGAGTGC 1622  
QY 3841 AGTGAATGCAAACTAGTCTCTCACGAAGAAGAGCTTGGCACCACTGCACAAATCCA 3900  
Db 1623 AGTGAATGCAAACTAGTCTCTCACGAAGAAGAGCTTGGCACCACTGCACAAATCCA 1682  
QY 3901 TTCTCTCAGATCCAGAAAGAAAGATATGTATCCATCCCATGCCACACCAAGCTGCTC 3960  
Db 1683 TTCTCTCAGATCCAGAAAGAAAGATATGTATCCATCCCATGCCACACCAAGCTGCTC 1742  
QY 3961 ACTTCGATATAGATTTGGGTAAAGAGATGATTTCCATCTGTTAGTAGGATCTATGAAT 4020  
Db 1743 ACTTCGATATAGATTTGGGTAAAGAGATGATTTCCATCTGTTAGTAGGATCTATGAAT 1802  
QY 4021 ATGGCTATGGCAGCTGGGA - ATGATAAAAATGGATCCAGATCTCAGCTTTAAACACAGAAG 4079  
Db 1803 ATGGCTATGGCAGCTGGCAATGTAAGTAATTAAGTCCAGATCTCAGCTTTAAACACAG - AG 1861  
QY 4080 ATTTTACCTGATGATCCAGACAAAGAACCCAGGCAAGAGCTACAGACCCGTCAGAG 4139  
Db 1862 ATTTTACCTGATGATCCAGACA - GAATCTCAAGGCAAAACAGCTACAGACCCGTCAGAG 1920  
QY 4140 TACCTCATTAATTTACTTGAATAAGACCTTGAAGAAAGAGTTCGCTAAAGGCTTCTG 4197  
Db 1921 TACCTCATTAATTTACTTGAATAAGACCTTGAAGAAAGAGTTCGCTAAAGGCTTCTG 1980  
QY 4198 GTCAGGCAATTTCCAAGAGAGGAACAAGAAATTAAGAAATTAAGATGAAGGCTTCAA 4257  
Db 1981 GTCAGGCAATTTCCAAGAGAGGAACAAGAAATTAAGAAATTAAGATGAAGGCTTCAA 2040  
QY 4258 AAATAAAGAAAGAAATTAAGAGTATTTCTCACCAACCCCTCAGAAAAATCTGATGAAG 4317  
Db 2041 AAATAAAGAAAGAAATTAAGAGTATTTCTCACCAACCCCTCAGAAAAATCTGATGAAG 2100  
QY 4318 ATGATGAGGAGGAGGATAACAAGTAAATGAAATGAAATCTGAAAAATAAGAAAAATCTA 4377  
Db 2101 ATGATGAGGAGGAGGATAACAAGTAAATGAAATGAAATCTGAAAAATAAGAAAAATCTA 2160  
QY 4378 AAAAAATTCATTTGCTGGATCTCCAGTTTCATATTTACTGCAACCCAGTGAACCCAGTTCTCA 4437  
Db 2161 AAAAAATTCATTTGCTGGATCTCCAGTTTCATATTTACTGCAACCCAGTGAACCCAGTTCTCA 2220  
QY 4438 TCTCAGAAATCTGAAGAACTCCATCAGAAAGACATTTAGTGTGCAAAAGAAAGATGA 4497  
Db 2221 TCTCAGAAATCTGAAGAACTCCATCAGAAAGACATTTAGTGTGCAAAAGAAAGATGA 2280  
QY 4498 GGCCTGTCAAAG 4509  
Db 2281 GGCCTGTCAAAG 2292





Db 2184 AACTGGCTAGCTCATCTCCCTGGTGCAGAAAATAATAGTGTAAATCCCTTGGTGAAGAAATGGGC 2243  
QY 1746 CTGGGTAAACAATAACAACAATTTCTTTCTGAACTTACCTGTGTTTCATGAACATCAACTG 1805  
Db 2244 CTAGGAAGACCATTCCAGACCATATCATCTCTCCTACCTGTCCACCAACACACAGCT 2303  
QY 1806 TATGGCCCTTTCTCTGCGCGTGCACCTTTCACCTTGACATCTTGGCGAAGAGAGATT 1865  
Db 2304 TATGGCCCTTTCTTATAGTCGTCCTTTATCCACCCTCACCTCATGGCAGAGAGATT 2363  
QY 1866 CAAACTTGGCTCCTCAGATGAATGCTGTAGTTTACTTAGAGATATACTACTAGAAAT 1925  
Db 2364 GAAATCGGCCACCCAGAGATTACGTAGTGTGTTTACATAGGTGACCTGATGAGCAGAAAT 2423  
QY 1926 ATGATAGGACTCATGAATGGATCCACACAGACTAAAGCATTAAGTTTAAACATACATT 1985  
Db 2424 ACGATAGGGAATATGAATGATTCATCCCAACCAAGATTTGAAGTTCAACGCACATT 2483  
QY 1986 CTGACGACATATGAATTTTACTGAAGGATPAAGTCAATCTCTGGTGTCTCAATTTGGCA 2045  
Db 2484 ATACACATATGAGATCTCTTTGAAAGATPAAGACTGTCTGGGCAGTATTAACCTGGGC 2543  
QY 2046 TTCATAGGAGTTCATGAAGCTCATCTGTTTAAATAATGATGACTCTCTCTGTACAGACT 2105  
Db 2544 TTTCTGGAGTGTGAAGCCCATCGGTGGAAGATGATGACTCTTTATTGTATAAACT 2603  
QY 2106 TTAATAGACTTTAAGTCCAACTCAGACTCTCTGATTACTGGAACCCCACTGCAAAATTC 2165  
Db 2604 CTGATTGATTTCAAGTCCAACTAGCTCTCTGATTACGGGACCCCTCTTCAGAAATTC 2663  
QY 2166 CTAAAGAGCTGTGCTTTGTTGTCATTCATCATGCGCAGAAAAATTTTCTCTCTGGGA 2225  
Db 2664 CTCAAAGAGCTCTGTCCTTCTGTCATTTATTATGCCGGAGAGTTTGAATTTTGGGA 2723  
QY 2226 GATTTTGAAGGAGATGCAAGGAAGAGATGTTGTTATCGAAGCTTTCACAAGAG 2285  
Db 2724 GATTTTGAAGACCATGGAAGGAGAGAGATTTGGCTACCAGAGTCTTTCATAAGGT 2783  
QY 2286 CTTTGAACCAATTTTACTAAGAAGAGTTTAAAGATGTAGAAAAGTCTTTACCTGCTAAG 2345  
Db 2784 CTAGAGCTTTCTCTCCGGAGAGTCMAAAAGATGTGGAGAAATCCCTTCTGCTAAA 2843  
QY 2346 GTTGACAAATTTCTAGGATGGAATGAGTGCATTTCGAGAGCAATATTACAAGTGGATT 2405  
Db 2844 GTGGAACAGATTTCTAGGGTGGAGATGTACAGCCCTTCAGAAACAGTATTACAAGTGGATT 2903  
QY 2406 TTAACAAGGAATTTATAAGCCCTCAGTAAAGTTTCAAAGGCGAGTACCTCAGGCTTTCTG 2465  
Db 2904 CTGACCAGGAATTTACAGGCTCTTGCCAAAGGAACAGAGGACACATCTGTTTTCCT 2963  
QY 2466 AACATTATGATGAAGTGTGTAAGAGTGTGTAACCATTTGCTACCTCATTTAAGCCACAGAT 2525  
Db 2964 AATATTGTGATGAACTGAAAAATGTGCAACCACTGCTATCTGATTAACCCCTGAA 3023  
QY 2526 GATAATGAATTTCTAATAACAGGAGGCTTACAGCATTTTGATAGTAGCAGCGGAAA 2585  
Db 3024 GAAATGAAAGGAAAATGACAGAGAGATTCTTCTCTCCCTCATTAAGGAGCAGTGGGAAG 3083  
QY 2586 CTAATCCTCTTGACAAGTACTGATTCGTCTCGGAGAACGTGGCAACAGAGTTCTGATT 2645  
Db 3084 TTGATTTTATAGCAAACTGTTGACNAGCTTCGAGAAGGGGAATCGAGTGTCTATC 3143  
QY 2646 TTTCTCAGATGTGAGGATGTGGACATCTCTAGCAGAAATATCTGAAGTATGCCAGTTT 2705  
Db 3144 TTTCTCAGATGTGAGAAATGTGGATATCTCTGGCTGAATACCTAACTATTAAACACTAT 3203  
QY 2706 CCCTCCAGAGACTTGATGATCAATAAAGGGGAATTTGAGGAGCAACCACTGGATCAT 2765  
Db 3204 CCTTTCCAGGCTGTGATGGTTCCATCAAGGGGAGAAATCCGAAAACAGGCACCTGGACCAC 3263  
QY 2766 TTCAATCGAAGGATCAGAGGATTTCTGTTTATTTTACTGTCTACAAGAGCTGGAGGATTA 2825  
Db 3264 TTCAATGCAGATGGTCTGAGGACTTCTGTTCTCTCTCGACAAAGGGCTGTGGCCCTG 3323

QY 2826 GSTATTAACTTGGCATCTGCTGACACTGTAGTTATTTTGAATCTGACTGAAATCCACAG 2885  
Db 3324 GGAATCAATTTGGCTTCACGGGACACAGTCGTCTATCTTTGACTCTGACTGGAACCCACAG 3383  
QY 2886 AATGATCTCGAGCCACAGGGAGAGCTCATAGAATTTGGACAGAGAACAAGGTTAATATT 2945  
Db 3384 AATGACTTTCAGAGCAACGCCGAGGCATAGAATTTGGTCAGAGAAGCAGGTAATATT 3443  
QY 2946 TATCGCTAGTGCACAAAAGGATCAGTAGAAGAAGATATCTTGAAGAGAGCAAGAAGAAG 3005  
Db 3444 TACCGCTTAGTTTACAAGGGGACTGTGGAGGAGAGATCATAGAACGGGCCAAAAGAAG 3503  
QY 3006 ATGGTCTAGACATTTAGTAAATTCAGAGAATGACACAGCAGCAAAAATCTGTTCTCAT 3065  
Db 3504 ATGTTATTAGATCATCTGGTGATTCAGCGCATGACACACCACTGCCCGGACGCTCTGAA 3563  
QY 3066 ACAGGTTTCAACTCCATCAAGCTCTACACCTTTTAAATAAGAAGATTATCAGCTATTTTG 3125  
Db 3564 AACAACTCAGGAAGGTCCAACCTCAAAATCTCTTTTAAATAAGAAGAGCTGACAGCTATT 3623  
QY 3126 AAGTTTGGTGTGAGGAACCTCTTTAAAGAACCTGAAGGAGAACAACAGGAGCCCAAGAA 3185  
Db 3624 AAATTTGGAGCAGAGGATCTCTTCAAAGAACTGGAAGGGAGGAATCAGAAACCTCAGAA 3683  
QY 3186 ATGGATATAGATGAAATCTTTGAAGAGAGCTGAAACCTCGGGAAAAATGAGCCAGTCCATTG 3245  
Db 3684 ATGGATATAGATGAAATTTTTCGGTTTGGCTGAAACAGAGAGAATGA--AGTGTCAACA 3740  
QY 3246 ACTGTAGGAGTACAGTTGCTTTCACAGTTCAGGTGGCGAACTTTTCCAATATGGATGAA 3305  
Db 3741 AGTGAACACAGATGAAGCTTCTATCACAGTTTAAAGTTTCCCAACTTTGCAACAATGGAGAT 3800  
QY 3306 GATGATATTGAGTTGGAACCCAGAAAGAAATTCAGAAATTTGGGAAGAAATCATCCAGAA 3365  
Db 3801 GAAGAA-----CAGCTAGAAGAGCTCTCAAGAGCTGGGATGAGATCATTTCCAGAG 3854  
QY 3366 TCCCAACGGAGAGGATAGAGGAGGAGAAAGAACAAAAAGAACTTTGAAGAAATATACAT 3425  
Db 3855 GAACAAAGGAAAAAAGTAGAGGAGGAGGCGCAGAGGAGCTAGAAGAAATTTATATG 3914  
QY 3426 CTCGCCAGGATGAGAACTGTGCAAAACAGATCAGCTTTAATGGGATGAA----- 3476  
Db 3915 CTGCCCTGGAATTCGGAGTTCCACTTAAAGGGCTCAGACAAATGACAGTACTGACACT 3974  
QY 3477 GGAAGACGAGTAGGACAGAGATATTCGGATCTGTAGTACTGACTC-----CATCACA 3530  
Db 3975 GAGTCTAAGAGGCGAGCCCGCAGAGATCCTCTGCTTCTGAGAGTGAACGGAGACTCTGAT 4034  
QY 3531 GAAAGAAAACGGCCAAAAGCGTGAAGACCTCGAACCATTCCTCGAGAAAAATATATAA 3590  
Db 4035 GATGACAAGAAGCAAGCGCAGAGGGCGCTCCGAGGAGTGTGCGGAAGGACCTCGTGGAG 4094  
QY 3591 GGATTTAGTGTGACAGATCAGCGGCTTATCAACAGTTTACAGAAATTTGGTGGCCCT 3650  
Db 4095 GGATTTACTGATCGAGAGATCCGAAGGTTTATCAAGGCTTATTAAGAGTTTGGTCTCCCT 4154  
QY 3651 CTGGAAGGTTAGATCTGTAGTAGATGCTGAACTGGTTGATATAATCTGAGACAGAC 3710  
Db 4155 CTGGAACGCTGGAGTGTAGCACGCTGATGCTGAGCTGCTAGTAACTGCGTGGCAGAT 4214  
QY 3711 CTTAGACGTTTGGTGAATTTGATATATATGAGTCAATTAAGGCTTTAAAGGACAATTC 3770  
Db 4215 CTGAAGCGCTGGTGAATGATCCACAACAGCTGTGTGTGTCAGCAATGCAAGGAATATGA 4274  
QY 3771 -----TCTGGCAAGAAAGAGCAGGAGTGTGGGAAGTTTAAAGCCCAACG 3821  
Db 4275 GAGCAGCTGAAAGAAATGCCGAGGAGAAAAGGACCAGGAGGAGAGGATCCAAACA 4334  
QY 3822 TTTTCGAATCTCAGGAGTGCAGGTGAATGCAAAACTAGTCTATCTCTCACGAAGAGAGCTG 3881  
Db 4335 ATCAAGATCCGAGTTCAGGTTTATGTGAATATCCCAATCCATATCCCAACATGAAGAGGATT 4394





```
QY 3663 GATGCTGTAGTAGAGTGTGAACCTGGTTGATAAATCTGAGACAGACCTTAGACGTTTG 3722
Db 721 GATGCTATACGTAGAGATGCTGAGCTAGTTGATAAATCTGAACAGACCTTAGACGCTG 780
QY 3723 GGTGAACCTGTACATATGATGATTAAGGCTTTAAAGACAAATTCATCTGGACAGAA 3782
Db 781 GGAGAACTTGTACATATGATGATTAAGGCTTTAAATGATATGACTTTGCTCAAGGA 840
QY 3783 AGAGCAGAGGTAGACTTTGGGAAAGTTAAAGGCCCAACGTTTCAATCTCAGCAGTGCAG 3842
Db 841 AGAACAGGTGTAGATTGGGAAAGTTAAAGGCCCAACATTCGATAGCAGGATGCAG 900
QY 3843 GTGAATGCAAACTAGTCTATCTCAGAGAAAGAGTGGCACCCTGACACAAATCCATT 3902
Db 901 GTGAATGCAAACTAGTCTATCTCAGAGAAAGAGTGGCACCCTATTCGATAAATCGATT 960
QY 3903 CCTTCAGATCCAGAAAGGAAAGATATGTATCCCATGCCACCAAGGCTGCTCAC 3962
Db 961 CCTTCAGATCCAGAAAGGAAAGATATGTATCCCATACCACCAAGCAGCTCAT 1020
QY 3963 TTCGATATAGATTGGGTAAGAGATGATTCCAATCTGTTAGTAGCATCTATGAATAT 4022
Db 1021 TTTGATATAGATTGGGTAAGAGATGATTCCAATCTGTTAGTAGCATCTATGAATAT 1080
QY 4023 GGTATGGCAGCTGGGAAATGATAAATGGATCCAGATCTCAGCTTAACACAGAAGATT 4082
Db 1081 GGTATGGCAGCTGGGAAATGATAAATGGATCCCTGATCTCAGTTTGACACAGAAGATT 1140
QY 4083 TTACCTGATATCCAGACAGAAACCCAGGCAAGCAGCTACAGACCCGTGAGACTAC 4142
Db 1141 TTACCTGATATCCAGATAAGAAACCCAGGCTAAGCAGTTACAGACTCGTGAGATTAC 1200
QY 4143 CTCATTAATTAATCTGATAAAGACCTTGCAGAAAGGAGCAACAAAGGCTTGCTGGTGA 4202
Db 1201 CTCATTAATTAATCTGATAAAGACCTTGCAGAAAGGAGCAACAGCAGACTTGTGTTGCA 1260
QY 4203 GGCAATTCAGAGAGAGGAGAGCAAGAAATPAGAGAAATAA 4243
Db 1261 GGCAATTCAGAGAGAGAGAAACAAAGAGTAAAGAAATAA 1301

RESULT 13
LOCUS AC019885 14530 bp DNA linear HTG 03-JAN-2000
DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
pieces.
ACCESSION AC019885
VERSION AC019885.1 GI:6665012
KEYWORDS HTG; HTGS_PHASE2.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 14530)
Adams,M. and Venter,J.C.
Direct Submission
Submitted (30-DEC-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
This sequence was identified as CDW:10211276 by the submitter.
For more information on this record e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
source Location/Qualifiers
1..14530
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
BASE COUNT 3407 a 3664 c 3613 g 3846 t
ORIGIN
```

```
Query Match 13.1%; Score 867.4; DB 2; Length 14530;
Best Local Similarity 55.0%; Pred. No. 4.2e-172;
Matches 2134; Conservative 0; Mismatches 1571; Indels 174; Gaps 15;

QY 1017 GAAGAAGATGAATTTGAAACTATAGAGAAGTTTATGGACATCGAATTTGCCCGAAAGGA 1076
Db 5569 GAGGAGGAAAGTGCAGAAACCATAGAGCGGTATCTCTGGCGCAGCGTGTGGAGAGAGGGGA 5510
QY 1077 GCACATGGTGCCCTCAACCAACCATCTATGCGGTTGAGCGAGATGSG--TGACCCCAATGCT 1133
Db 5509 TGCACGGGCAATCAGACAACGATCTATGCAATAGAGAAAACGGCTTTGATCCACATGCT 5450
QY 1134 GGGTTTG-----AAAAGTCAAAAGGAGCTGGGAGAAATACAGTATCTTATTAAATGG 1184
Db 5449 GGATTTGATGAAAGCAAACTCCAGACGGGAGACCGGACAGTTCCTTGATCAAGTGG 5390
QY 1185 AAGGCTGGTCAACATATCAATACACTTTGGGAACTTGGAAGAAACGCTGGAAGCAACAAAT 1244
Db 5389 AAGGCTGGTGGTACATCCACAACACTTGGGAATCTGAGGCTACCCCTGCGTGATGAAG 5330
QY 1245 GTTAAAGGAATGAACAAACTTGGCAACTACAAGAAAAGGATCAGGAGACAAAACGCTGG 1304
Db 5329 GCCAAGGAATGAAGAAGCTGGATACTTCAATCAAAAAGGAGAAAGCAAGGCTACTGG 5270
QY 1305 CTGAAAATGCTTCTCCAGAAGATGTGGAATATTATACTGCCAGCAGGAGCTTACAGAT 1364
Db 5269 CGCCGATACGCTGGCCCCGAGGACATCGACTACTTCCAGTGTGAGCTGGAACCTGACAT 5210
QY 1365 GATCTGCACAAACATATCAATAGTGGAAAGAAATTAATTTGCTCATTTCAATCAAAAGTCA 1424
Db 5209 GAGCTACTGAAATCGTACAAACACGCTGGACCGAATCATTTGCCAAGGGCTCT---AAGCCA 5153
QY 1425 GCAGCTGGTTATCCGACACTACTTCAATGGCAGGCTCTGCTTACTCAGAAATGTAGC 1484
Db 5152 GAGCAGCGCACCGAGAGTACCTGTGCAATATGGCAGCTCGTACCGAGTCCACACA 5093
QY 1485 TGGGAAGATGGTGCTCTCATTTGCCAAAAGTTTTCAGGACGCAATTTGATGAGTATTTAGC 1544
Db 5092 TGGGAGGATGCCGCCCTGGTTCTTCGCAAGTGGCAACGCTGCGCAGAGCAATTTCAATGAC 5033
QY 1545 AGAAATCAATCCAGAGACTACTCCCTTTAAGGACTGAAGGTTCTTAAACAGACCAAGA 1604
Db 5032 CGGAGTCTCTCAAAATGCACTCCCTCCGCACTGAGAGTGATCAAGTATCCGCCCAAG 4973
QY 1605 TTTGTTGCACTGAAGAAGCAACCATCTTACATTTGGAGACATGAAAGTCTTGGAGTTAAGA 1664
Db 4972 TTCTCTCGAATAAAGATCAGCCGGAATTTCTTCTGTCGGCCCTAACTC-----TAAGG 4919
QY 1665 GATTATCAGTTAAATGGATTGGCTGCGCTCATTTATGGTGCAAGAAATAGTTGT 1724
Db 4918 GATTACAAATGGACGGCCTGAAATTTGGCTGCTGCACTCATGTTGCAAGGAGAACTCGGTG 4859
QY 1725 ATCTTTGCAGATCAAAATGGTCTGGGTAAACAATACAACAATTTCTTTCTTGAACTAC 1784
Db 4858 ATCTGGCCGACGAGATGGCCCTCGGCAAGACCATCCAAACTATTTGCTTCTGTACTCG 4799
QY 1785 CTGTTTCATGAACATCAACTGTATGGCCCTTTTCTTCTGCGCTGCCACTTTTCTACTCTG 1844
Db 4798 CTGTTCAAGATCCATCACTGTACGCCCTTCTCTGCGCTGCTGCACTTAGCACAATG 4739
QY 1845 ACATCTTGGCAAGAGAGATTCAAACTTTGGGCTCTCAGATGAATGCTGTAGTTTACTTA 1904
Db 4738 ACGGCTTGGCAGCGGAGTTCGACCTTGTGGCGCCGAGACATGTTAGTTCAGCTACCTC 4679
QY 1905 GGAGATATACTAGTAGAATATGATAAGCACTCAATCAATGATGATGCATCCACAGACTAAA 1964
Db 4678 GGTGACATTAATCCCGGAGCTCATTTCAACAGTAGAATGGCAGTTTGAGAGCTCCAAG 4619
QY 1965 CGATTAAAGTTTAACATATCTTGACGACATATGAATTTTACTGAAGGATAAGTCAATTC 2024
Db 4618 CGGCTTAATTCATTTGCATTTCTCACACGCTACGAGATTTGCTCAAGGATAAACAATTT 4559
QY 2025 CTTGGTGGTCTCAATTTGGGCATTTCAAGGAGTTGATGAGAGCTCATCGTTTAAAAAATGAT 2084
```

Db 4558 CTGGGACGGCTTCAGTGGGCTGGCTCTCTGGTGGACGAAGCGATCGCTTTAAGAACGAC 4499  
Qy 2085 GACTCTCTCTGTACAGGACTTTAATAGACTTTAAGTCCCAACCATCGACTTCTGATTACT 2144  
Db 4498 GACTCACTGCTGTACAACTGCTGAAGGAGTTTGACACATAACCATCGACTCTCATAA 4439  
Qy 2145 GGAACCCCACTGCAAAATTCCTCAAGAGCTGTGGTCTTTGTTGCTATTCATCATGCCA 2204  
Db 4438 GGTACCCCGCTGCAGATTCGCTGAAGGAGCTCTGGCTTTGCTGCACTTCATCATGCCG 4379  
Qy 2205 GAAAAATTTCTCTCTGGGAGATTTTGAAGGAGCATGCAAGGAGAGAGATGTGT 2264  
Db 4378 GATAAGTTTGACACCTGGGAGATTTTCAGGTGCGAGCATGCGCAATCGCGGAGCAAGGC 4319  
Qy 2265 TATCAAGTCTTCAACAAGAGCTTGACCATTTTACTAAGAGAGTTTAAAAAGATGA 2324  
Db 4318 TACACTCGGTGCATCAACAGCTGGAGCGGTATTTCTGCGAGAGTGAAAGAGATGTG 4259  
Qy 2325 GAAAGTCTTTACCTGCTAGGTTGAACAAATCTGAGGATGGAATGAGTGCATTCGAG 2384  
Db 4258 GAGAAATCGCTGCCGCTTAAAGTGGAGCAGATCCTGCGGGTCGAGTGACCTCTCTGCGAG 4199  
Qy 2385 AAGCAATATTACAAGTGGATTTTAAACAGGAAATATAAGCCCTCAGTAAAGGTTCAAA 2444  
Db 4198 AAACAGTACTACAAGTGGATCTTAACCAAAAACCTTTGATGCTGCGCAAGGAAAGCGG 4139  
Qy 2445 GCGAGTACCTCAGGCTTTCTGAACATTTATGATGGAACCTTAAAGAGTGTGTAACCATTCG 2504  
Db 4138 GCGAGTACCTCCACTTTCTTCAACATAGTCATCGAGCTTAAAGAGTGTGTCACACGCG 4079  
Qy 2505 TACCTCATTAAGCCACCAAGATGATTAATGAATTTCTATATAAACAG ---GAGGCCCTTACAG 2561  
Db 4078 GCTCTCATCGACCATCCGAGTTTGAGCTTTAGGCTCTGACGACGATGAGGCATTGCA 4019  
Qy 2562 CATTGTATGATCGACGCGGAACCTAATCTCTTTCAGACGACTGATTCGTCTGCA 2621  
Db 4018 ACTCTTTTAAAGGATCCGAAAGCTTGTGCTGCTGATAGTACTCTGCGGACTGAA 3959  
Qy 2622 GAACGTGGCAACAGATTCGATTTCTCTAGATGTTGAGAGTGTGGACATCCTTAGCA 2681  
Db 3958 GAGAGGGCCACCGGCTGATTTTCCCAATAGTGGCGCATGTGACGCTTGTGGCT 3899  
Qy 2682 GAATATCTGAATTCGCCAGTTTCCCTTCCAGAGACTTGATGATCAATAAAGGGGAA 2741  
Db 3898 GACTATCTGCAAGCGCTCACTTCCCTTTCCAGCGCCTCGACGCGAGCATAAAGGGAGAG 3839  
Qy 2742 TTGAGGAAGCAAGCACTGGATTCATTTCAATGCAAGAGATCAGAGGATTTCTGTTTTTA 2801  
Db 3838 ATGAGGCGCGAGGCTCTGGATCACTTCAATGCAAGAGGATAGCCAGGACTTTTGTCTCTG 3779  
Qy 2802 CTGCTTCAAGAGCTGGAGATTAGGTATTAATCTTGGCATCTGTCACACTGTAGTTAT 2861  
Db 3778 CTTTGCAAGAGGCTGTGGATTGGGATTAATCTTGGCCACAGCCGACTTTTGTCTCTG 3719  
Qy 2862 TTTGATTCTGACTGGAATCCACAGAATGATTCGACGACGAGGAGGACTCATAGAAAT 2921  
Db 3718 TTTGACTCCGACTGGAACCCCAACAGATTTTACAGGCGCAGCGAGAGCGCATCGATC 3659  
Qy 2922 GGACAGAAGAAACAGTTAATATTATCGCTGAGTCAAAAGGATCAGTAGAGAGAT 2981  
Db 3658 GGTCAAGAAAACAGTTTACATTTATCCCTGTGTCACCCCGCTGCTGGTGGAGAACAG 3599  
Qy 2982 ATTTCTTGAAGAGCAAGAGAGATGTTGCTAGACCATTTAGTAAATTCAGAGATGGAC 3041  
Db 3598 ATCGTGGCGGAGCAACAGAAATGGTACTTGGACCACTGCTCATTCAGCGGATGGAC 3539  
Qy 3042 ACGACAGGAAAACCTGTTCTGCATACAGGTTTCAA -----CTCCATCAAGCTCTACACCT 3095  
Db 3538 ACTACAGGAGGACACTGTTCTGGATAAGCGCGCAACGGGCACTCTGCTTAACCTCGAATCCG 3479  
Qy 3096 TTTAATAAGAGAGTTATCAGCTATTTTGAAGTTTGGTCTGAGGAACTCTTTTAAAGAA 3155

Db 3478 TTCAACAGGACGATCTTTCTTCTGCTATCTCTGAAGTTTGGTGCAGAGGAGCTGTTCAGGA - 3418  
Qy 3156 CTTGAAGGAGAAACAGAGAGCCCGAGAAATGGATATAGATGAAATCTTTGAAGAGACT 3215  
Db 3419 --CGAACAGAGCATGACGAGGACTTGGTTTGTGACATTTGATGAAATCTTCGCGAGCGC 3362  
Qy 3216 GAAACTCGGAAATGAGCCAGGTCCATTGACTG -----TAGG 3254  
Db 3361 GAGACCGCAACGAAGATCCGAAATGCCAGCAGACGACTTGTGTGCTCCCTTCAAAGTG 3302  
Qy 3255 GATGAGTTGCTTTCACAGTTCAAGTGGCGAACTTTTCCATATATGATGATCAAGATGATT 3314  
Db 3301 GCCAGCTACTCTCTTTCGAGAGAAACCGAGGATTCGGTTAGCAAGCAACCAAAAT 3242  
Qy 3315 GAGTTGGNACCAAGAAATTTCAAGAAATTTGGGAATTCATCCCAAGATCCCAAGG 3374  
Db 3241 GCCCGCGGAGGAGTACAGTAGGACTGGGACACATCTTCGGAAGGCTTCGT 3182  
Qy 3375 AGAAGATAGAGAGAGGAAAGACAAAGAAACTTTGAAGAAATATACATGCTCCGAGG 3434  
Db 3181 AAAGCAATTCAGCATCAGGAGCGAGCAAAAGAGATGGAAGATCTCTACTTGGCCACCGG 3122  
Qy 3435 ATGAGAACTGTGCAAAACAGATCAGCTTTTAAATGGGAGTGAAGAGACGCACTAGTAG - 3493  
Db 3121 AGGAAACTGCCCTTAACCAAGGAAGGAGCGTGGAGCTGGCAAGGGGGGCAAGGG 3062  
Qy 3494 -----CAGAGATATTCTGGATCTGATAGTACTGACTCCATCACAGAA 3533  
Db 3061 AAACAGGACGACGACTCCGCTGGCGACTCGGATTACGAGCTGGGCTCCGCGGCACT 3002  
Qy 3534 AGAAACGGCAAAAGCGTGGAGACCTCGAACCATTCCTCGAGAAATATTAAGGA 3593  
Db 3001 GGTGAGCTGGACCTCGCAAGCGCGCGCGCCGACCATGAGGAGAGATCACCGA 2942  
Qy 3594 TTTAGTATGACAGATCAGCGGTTTATCAAGAGTTACAGAAATTTGGTGGCCCTCG 3653  
Db 2941 TTCATGATGCGAGTTGCTGCTCTCTCCGAGTTACAAAAGTTTCCCTCTCTCTC 2882  
Qy 3654 GAAGGTTAGATCTGAGTAGAGATGCTGAACCTGGTGGTGAATCTGAGACAGACTT 3713  
Db 2881 CACCGATGAGGCGCATCGATGATGCTGAGCTGCAGAAAGCCACTTGGCGGAGCTG 2822  
Qy 3714 AGACGTTGGGTGAATTTGTATATAATGAGATTAAGGCTTTTAAAGGACAATTCATCT 3773  
Db 2821 AAGCGCTCGAGAGATGCTCATGACCGCTGCTTCAAGTTCTTCCACAGCAGCAAGAG 2762  
Qy 3774 GGACAAGAAAGACGAGGATGAGCTTTGGGAAAGTTTAAAGGCCCAACGTTTTCGAA ----- 3828  
Db 2761 GAAGAGATGAGATCGCGCGACGATGAGACCGCGCGCCCAAGCAGCGCGCGCAGC 2702  
Qy 3829 -----TCTCAGGAGTGCAGTGAATGCMAAATCTAGTCTCTCTAC 3869  
Db 2701 GCCACTTCTCGTTAAGCTGGTGGCTCTCTCTTAAATGCCAAAAGCTGTGCGCTGC 2642  
Qy 3870 GAAGAAGAGCTGCGACCTACCAAAATCCATTCCTTCAGATCCAGAAAGAAAGAA 3929  
Db 2641 GAACAGGCTTCAGCGCTCAAGAGATCATGCCGACATGCGCGAGAGCGCCAGCA 2582  
Qy 3930 TATGTCATCCATGCGCACACCAAGCTGCTCACTTCGATATAGATGGGTAAAGAGT 3989  
Db 2581 TGGAGCTTCAATATCAAGNACAGCGCGCGCTTTCGAGCTGATTTGGGCGCATCGAGAG 2522  
Qy 3990 GATTCCAATCTGTAGTAGGCACTATGAAATGCTATGGCAGCTGGGAAATGATAAA 4049  
Db 2521 GACAGAGAGCTTACTGCGGCAATTTACCAGTACGCGATTTGATCTGGGAGCAGATGAG 2462  
Qy 4050 ATGGATCCAGATCTCAGCTTAAACACAGAGATTTTACCTGATGATCCAGCAAGAAACC 4109  
Db 2461 CTAGACCCACGCTCAACTACCCGACAGATTTTGTCTTAAAGATACGCGCAAG ---CCC 2405  
Qy 4110 CAGGCAAGCAGCTCAGACCCGCTGACACTACCTCAATTAATTTACTGATTAAGACCTT 4169  
Db 2404 CAGGCAAGCAGCTGACAGCGCTGCGGAGTACCTGCTCAAGATCATCAAGAAGAGCTG 2345











QY 2982 ATTCTTTGAAGAGCCCAAGAGATGGTGTAGACCAATTTAGTAATTCAGAGAAATGGAC 3041  
Db 69199 ATCGTGAGCGGCCAAACACAGAAATGGTACTGGACCACTGTCTCATTCAGCGGATGGAC 69258  
QY 3042 ACAGACAGGAAAATGTTCTGCAATACAGGTTCAA-----CTCCATCAGAGCTCTACACCT 3095  
Db 69259 ACTACAGAGCGCACTGTTCTGGATAAGAGCGGCAAGCGGCACCTCGTCTAACTCGAATCCG 69318  
QY 3096 TTTTAATAAAGACAGATATACAGCTATTTGAAGTTTGGTGTGCTCAGCAATCTTTAAAGAA 3155  
Db 69319 TTAACAAGGACCATCTTCTGCTAATCTGGAAGTTGGTGAGAGAGCTGTTTCAAGGA- 69377  
QY 3156 CTTGAAGGAGAAAGACAGAGCCCGGAAATGGATATAGATGAATCTTTGAAGAGAGCT 3215  
Db 69378 -CGACAGAGCATGACGACGACTTGGTTGTGACATTTGATGAATCTTGCOCGAGGCG 69435  
QY 3216 GAAACTCGGAAAATGAGCGAGCTCCATTTGACTG-----TAGGG 3254  
Db 69436 GAGACCCGCAACGAAGATCCGGAATGCCAGCAGAGCACTTGTGTCCGCTTTCAAGTG 69495  
QY 3255 GATGAGTTGCTTTACAGATTCAAGGTGGGAACTTTTCCAAATATGGATGAAGATGATTT 3314  
Db 69496 GCCAGCATAGCTCTTTCGAAGAGGAAACCGAGCGATTCGGTTAGCAAGAACCAAAAT 69555  
QY 3315 GAGTTGGAACCAAGAAATTTCAAGAAATTTGGGAAGAAATCATCCAGAAATCCCAACGG 3374  
Db 69556 GCCCGCGGAGGAGTACAGTAAGGACTGGGACGACATTCCTCGGAAGCTTTCCGT 69615  
QY 3375 AGAAGATAGAGAGGAGGAAAGACAAAGAACTTTGAAGAAATATACATGCTCCCGAGG 3434  
Db 69616 AAGCAATCGACATCAGGAGCGGAGGAAAGAGATGGAAGATCTACTTGGCACCCCGG 69675  
QY 3435 ATGAAACTGTCGCAAAACAGATCAGCTTTAATGGGAGTGAAGGAAGAGCAGTATGAG- 3493  
Db 69676 AGGAAACTCCCTTACCAAAACGAGGAAACGCTGGAGCTGCAAGGGGGCAAGGG 69735  
QY 3494 -----CAGAAGATATCTGGATCTGATAGTACTGACTCCATCAGAA 3533  
Db 69736 AAACAGAGCAGCAGACTCCGCTGGCGACTCGGATTAACGAGCTGGCTCCGACGCACT 69795  
QY 3534 AGAAACGGCCAAAAGCGTGAAGACCTCGAACCATTCCTCGAGAAATATTAAGGA 3593  
Db 69796 GGTGAGATGGAGCCTCGCAAGCGCGGCGCCACCACCATGAAGGAGAGATCAACCGGA 69855  
QY 3594 TTTAGTGCAGAGATCAGCGGCTTATCAAGAGTTTACAAGAAATTTGGTGCCCTCTG 3653  
Db 69856 TTTACTGATGCGAGTTCGCTCCTTCATTCGCAAGTTTACAAAAGTTTCCGCTCCTC 69915  
QY 3654 GAAAGTTAGATGCTGTAGTAGATGCTGAACTGGTTGATAAATCTGAGACAGACCTT 3713  
Db 69916 CACCGATGGAGGCCATCGCATGTGATGCTGAGCTGCAGGAAAAGCACTGGCGGAATG 69975  
QY 3714 AGAGTTTGGTCAACTGTACATATGATGATTAAGCTTTTAAAGCAATTCATCT 3773  
Db 69976 AAGCGCTCGGAGATGCTGCAATGACCGCTGGTTGAGTTCTGACGAGCAAGGAG 70035  
QY 3774 GGACAAAGAGACAGGAGTAGACTTGGGAAAGTTTAAAGGCCCAACGTTTTCGA----- 3828  
Db 70036 GAAGAGATGAAGACTGGGCGAGGATGAGACCGCGGCGCAAGCAGCGCCGCGCAGC 70095  
QY 3829 -----TCTCAGGAGTGCAGGTGAATGCAAAACTAGTATCTCTAC 3869  
Db 70096 GCCACCTTCTCCGTTAAGCTGGTGGCTCTCCTTTAATGCCAAAAGCTGTGGCTGC 70155  
QY 3870 GAAGAGAGCTGGCACCACTGCAACAAATCCATTCCTTCAGATCCAGAAAGAAAGA 3929  
Db 70156 GAACAGAGCTTCAGCGCTCAACAGAGATCATGCCAGCATGCCGAGAGGCCAGCAA 70215  
QY 3930 TATGTCATCCATGCCACACCAAGGCTGCTCACTTCGATATAGATTGGGTAAAGAAAGAT 3989  
Db 70216 TGGAGCTTCAATATCAAGACACCGCCCGGCTCTCGACGTGATTGGGCAATCGAGGAG 70275  
QY 3990 GATTCCAATCTGTAGTAGGCATCTATGAATATGGCTATGGCAGCTGGGAAATGATAAAA 4049

Db 70276 GACACGAAGCTACTGTGGGCAATTTACCAGTACGGCATTTGGATCCTGGGAGCAGATGAAG 70335  
QY 4050 ATGGATCCAGATCTCAGCTTTAACACAGAGATTTTACCTGATGATCCAGACAGAAACCC 4109  
Db 70336 CTAGACCCCAAGCTCAAACTGACCGCAAGATTTTGTCTAAACGATACGCGCAAG---CCC 70392  
QY 4110 CAGGCAAGCAGCTACAGACCCGTCAGACTACCTCATTAATTTACTTGAATAAGACCTT 4169  
Db 70393 CAGGCCAAGCAGCTGACAGCGCTGCCGAGTACTGCTCAAGATCATCAAGAAAGACGTG 70452  
QY 4170 G----- 4170  
Db 70453 GAGCTGACCAAGGAGGACAAACCGCGACACGCGCCCGCAGAGCATCGCGAGCTAAGCAT 70512  
QY 4171 --CAAGAAAGGAAGACAAAGGCTTGTGTGAGCAATTTCCAAAGAGAGGAGACAAG 4228  
Db 70513 GCCAAGCGGCAAGCCAGTGTCTAGTTGACCATGATGCTTAAGCGCGCAGATGGCGAG 70572  
QY 4229 AATAAGAGAATAAGATGAAGGCTTCAMAAATAAAGAAAGAAATAAAGAGTGTATCTTC 4288  
Db 70573 GAGCTGTGGCAGCGACCGCACCTGCGCGAAAGCAGTAACAGTCAAGTAGATCCATCT 70632  
QY 4289 ACCACAAACCTCAGAAA-----AATCTGATGAAGATGATGAGGAGGAGGATACAAAGGTA 4343  
Db 70633 ACCGCTGCGCGCACAAATGCTCCAGCTACGGAACAGCACGCTGACCTGCGAAGAAAGCC 70692  
QY 4344 AATGAATGAATCTCAAAATAAAGAAATCTAATAAATAATTCATTTGCTGGTACTCCA 4403  
Db 70693 AAAAAATCCAAAGCTCTGCGAAGAAACAGCGCTCGGACAAACACGCGCAACAGCCG 70752  
QY 4404 GTTCATATTACTCAACCAAGTGAACAGTTCTCTATCTCAGAAGAAATCTGAAGAACTCCAT 4463  
Db 70753 ATGCATTTACGCGCAACATGAACCTAGGGCTTT---GGAGTCTCTGGTGTACCTAGAT 70809  
QY 4464 CAGAAGACATTTAGTGTGTCTCAAGAAAGAAATGAGGCTGTCTCAAGCAGCACTGAAACAG 4523  
Db 70810 CCCAGCATTTTAAAGAGTCAAAAGAAATAATCGCGCGGTGAAGAGGCGCTCAAGGCG 70869  
QY 4524 CTGGATAGACCAAGAGAGGCGCTTCTGAAGGAGCAGCTGGAACATACTAGCAGTGT 4583  
Db 70870 CTGGACCGCGGATGTCTGCTTCCGACAGGATCAGCTGCAGCAGCAGGAGGATTGC 70929  
QY 4584 CTAATCAAAATTTGGGATCACATTACAGATGCTCTGAAGAGGTACACAAATCCCGAGCAA 4643  
Db 70930 CTGCTTCAATTTGGCAAGCAGATCGATGTTGCTCAATCCCTACCGCGAGACAGAG--- 70986  
QY 4644 ATAAACAGTGGAGGAAAAATTTGTGATTTTGTGTCTCAAGTTTACAGAAATTTGATGCC 4703  
Db 70987 AAAAAAGAAATGGCGCAGTAACTTTGTTGTTGTCGAAATTTACTGAGCTGGATGCC 71046  
QY 4704 AGAAAGCTGCACAACTCTACAAACATGCAATCAAAAAG 4742  
Db 71047 AAGCGCTATTCAAGATTTTACAAAACAGCACTCAAGCAG 71085

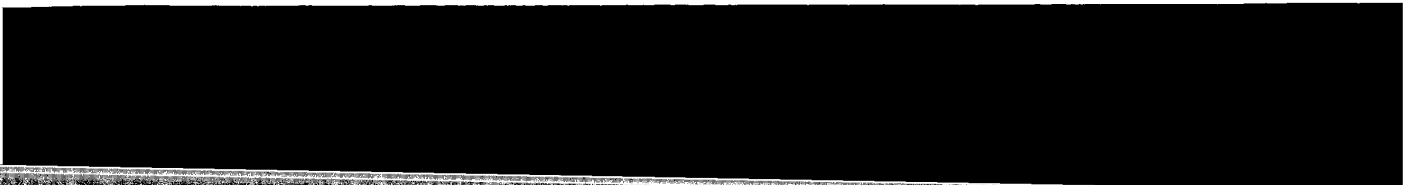
Search completed: August 2, 2002, 20:55:27  
Job time: 26568 sec



Mon Aug 5 11:51:36 2002

us-08-973-363-10.rge

Page 35



\_\_\_\_\_

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 2, 2002, 20:47:37 ; Search time 1205.34 Seconds  
(without alignments)  
9412.590 Million cell updates/sec

Title: US-08-973-363-10

Perfect score: 6608

Sequence: 1 CGGGCTGGCGACGAGGCGC.....AACTAAACCAAAAAAACCC 6608

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N\_Geneseq\_032802:\*

1:	/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1980.DAT.*
2:	/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1981.DAT.*
3:	/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1982.DAT.*
4:	/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1983.DAT.*
5:	/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1984.DAT.*
6:	/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1985.DAT.*
7:	/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1986.DAT.*
8:	/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1987.DAT.*
9:	/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1988.DAT.*
10:	/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1989.DAT.*
11:	/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1990.DAT.*
12:	/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1991.DAT.*
13:	/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1992.DAT.*
14:	/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1993.DAT.*
15:	/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1994.DAT.*
16:	/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1995.DAT.*
17:	/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1996.DAT.*
18:	/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1997.DAT.*
19:	/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT.*
20:	/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT.*
21:	/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT.*
22:	/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT.*
23:	/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT.*
24:	/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	ID	Description
1	6608	100.0	6608 18 AAT42751	Chicken CHD-1A gene
2	1098.2	16.6	1316 18 AAT42754	Chicken CHD-W gene
3	867.4	13.1	6240 23 ABL06453	Drosophila melanog
4	867.4	13.1	9933 23 ABL06442	Drosophila melanog
5	553.2	8.4	4222 23 AAS85967	DNA encoding novel
6	321.4	4.9	575 21 AA280469	Human colon cancer
7	311.2	4.7	6475 22 AAF58444	Human polyomucleoti
8	309.6	4.7	6327 17 AAT32301	Dermatoyomitis sp
9	296.6	4.5	4177 22 AAD02817	Arabidopsis thalia

10	274.2	4.1	3031	23	AAS73906	DNA encoding novel
11	266.8	4.0	8112	23	ABL06515	Drosophila melanog
12	266.8	4.0	10274	23	ABL06514	Drosophila melanog
13	245.8	3.7	3700	23	ABL14275	Drosophila melanog
14	239.4	3.6	6553	22	AAI60230	Human polyomucleoti
15	236.6	3.6	2195	22	AAH15817	Human cDNA sequenc
16	230.4	3.5	7545	23	AAS73907	DNA encoding novel
17	230.4	3.5	8617	22	AAK52052	Human polyomucleoti
18	226.4	3.4	3054	23	AA574043	DNA encoding novel
19	221	3.3	2512	22	AAH14191	Human cDNA sequenc
20	221	3.3	3901	22	AA561042	Human cancer agent
21	206.6	3.1	6011	23	ABL14274	Drosophila melanog
22	191.6	2.9	2936	22	AAH17827	Human cDNA sequenc
23	187.8	2.8	4969	23	ABL20345	Drosophila melanog
24	187.8	2.8	5553	23	ABL10259	Drosophila melanog
25	174	2.6	3219	21	AAF16302	Human prostate can
26	173.6	2.6	672	22	AA540963	DNA encoding nove
27	172	2.6	971	22	AA531113	Human diagnostic a
28	168.2	2.5	806	22	AA197310	Human neuroblastom
29	165.4	2.5	1425	21	AA577085	Human ORFX ORF2640
30	162.8	2.5	3386	21	AA538904	Human tumour suppr
31	162.8	2.5	5468	21	AA538907	Human tumour suppr
32	162.8	2.5	5471	21	AA538846	Human tumour suppr
33	162.8	2.5	5471	21	AA538905	Human tumour suppr
34	162.8	2.5	5477	21	AA538911	Human tumour suppr
35	162.8	2.5	5480	21	AA538909	Human tumour suppr
36	162.8	2.5	5564	21	AA538908	Human tumour suppr
37	162.8	2.5	5567	21	AA538906	Human tumour suppr
38	162.8	2.5	5573	21	AA538912	Human tumour suppr
39	162.8	2.5	5576	21	AA538910	Human tumour suppr
40	159.4	2.4	1365	22	AA563216	Human purified sec
41	156.6	2.4	3436	22	AA580550	Receptor #38 parti
42	155.6	2.4	964	22	AA540966	cDNA encoding nove
43	151.4	2.3	337	14	AA060122	Human brain Expres
44	149	2.3	1311	19	AAV59280	Altered telomere r
45	148	2.2	5994	21	AA70222	Plasmodium falcipa

## ALIGNMENTS

RESULT 1

AAT42751  
ID AAT42751 standard; cDNA: 6608 BP.

AC AAT42751;

DT 12-MAR-1997 (first entry)

DE Chicken CHD-1A gene.

FW Bird; sex determination; chromodomain-Helicase-DNA binding 1 Avian;

CHD-W; chromodomain-Helicase-DNA binding on the W chromosome; ss.

OS Gallus sp.

Key Location/Qualifiers

FT CDS 228..5390

FT /\*tag= a

PN WO9639505-A1.

PD 12-DEC-1996.

PF 05-JUN-1996; 96WO-GB01341.

PR 06-JUN-1995; 95GB-0011439.

PA (ISIS-) ISIS INNOVATION LTD.

PI Griffiths R, Tiwari B;

DR WPI; 1997-043127/04.









Db	901	gtgaatgcaaacgctagtcattctccaggaagaagtggtggccaccattgcataaatogatt	960
Qy	3903	CCTTCAGATCCAGAAAGAAAGAAAAGATATGTCATCCCATGCCACACCAAGGCTGCTCAC	3962
Db	961	cttcagatccagagaagaagaagatatgtcatcccataccacacccaagcagctcat	1020
Qy	3963	TTTCGATATAGATTGGGGTAAAGAAGATGATTCCAATCTGTTAGTAGGCACTCTATGAATAT	4022
Db	1021	tttgatagattgggttaagaagatgattccaatctgttaataggcactctatgaatat	1080
Qy	4023	GGCTATGGCAGCTGGGAATGATAAAATGGATCCAGATCTCAGCTTAAACACAGAAGATT	4082
Db	1081	gggtatggcagttgggaaatgataaaaatggatcctgctcagttctgacacagaagatt	1140
Qy	4083	TTACCTGTATGATCCAGACAAGAAACCCACAGGCCAAAGCAGCTACAGACCCGTGCAGACTAC	4142
Db	1141	ttaccctgagatccagataagaaccccaaggctaaagcagctctacagactctgcagattac	1200
Qy	4143	CTCATTTAAATTTACTGTAATAAGACCTTCACAAGAAAGGACACAAAGGCTTGCTGGTGCA	4202
Db	1201	ctcataaattactgaataaagacaccttgcagaaaggaagcacagagacttgcgtggtgca	1260
Qy	4203	GGCAATTCAGAGAGGAGGACACAGAAATAGAGAGAATAA	4243
Db	1261	ggcaattcaagaggagaaacaaagaagtaagaagataa	1301
RESULT 3			
ABL06443			
ID	ABL06443	standard; cDNA; 6240 BP.	
AC	ABL06443;		
CC	26-MAR-2002	(first entry)	
DE	Drosophila melanogaster expressed polynucleotide SEQ ID NO 13811.		
KW	Drosophila; developmental biology; cell signalling; insecticide;		
KW	pharmaceutical; gene; ss.		
OS	Drosophila melanogaster.		
PN	WO200171042-A2.		
PD	27-SEP-2001.		
PF	23-MAR-2001; 2001WO-US09231.		
PR	23-MAR-2000; 2000US-191637P.		
PR	11-JUL-2000; 2000US-0614150.		
PA	(PEKE ) PE CORP NY.		
PI	Venter JC, Adams M, Li PWD, Myers EW;		
DR	WPI; 2001-656860/75.		
DR	P-PSDB; ABB62340.		
PT	New isolated nucleic acid detection reagent for detecting 1000 or more		
PT	genes from Drosophila and for elucidating cell signalling and cell-cell		
PT	interactions -		
PS	Claim 1; SEQ ID NO 13811; 21pp + Sequence Listing; English.		
CC	The invention relates to an isolated nucleic acid detection reagent		
CC	capable of detecting 1000 or more genes from Drosophila. The invention		
CC	useful in developmental biology and in elucidating cell signalling and		
CC	cell-cell interactions in higher eukaryotes for the development of		
CC	insecticides, therapeutics and pharmaceutical drugs. The invention		
CC	discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA		
CC	sequences (ABL01840-ABL16175) and the encoded proteins		
CC	(ABB57737-ABB72072).		
CC	The sequence data for this patent did not form part of the printed		











Db 3710 AGATCTAAGTCAAAAATGGAAGAAGATTCTTGACAAAAAAGAGACAGATGATTC 3651  
 QY 867 TC----AGAGGAGGAGGACGATGATGAAGATTATGATAAGAGAGATCTGTCGCCA 922  
 Db 3650 TCTGAGAGGATGATGATGAAGAAGATTATGATAAAGAAAGTCTGTCGCCA 3591  
 QY 923 GGCACAGTGAATGTTAGTTACAAAGAAGCTGAAGAACCAAGACAGATCTGATGATTT 982  
 Db 3590 AGCAACTGTTAATGTTAGCTATAAGGAGGATGAAGAAATCAAAACAGATCTGATGACCT 3531  
 QY 983 GCTGGAAGTTTGTGAGAGGATGTCACAGACTGAAGAAGATGAATTTGAAACTATAGA 1042  
 Db 3530 ACTGGAAGTCTGGAGAGGATGTTCTCAACTGGAGAGAGGATTTGAAACCATAGA 3471  
 QY 1043 GAAGTCTTATGACAGTCCGAATTTGGCGGAAGAGGACCACTGGTCCCTCAACCATCTA 1102  
 Db 3470 AGATTTATGGATTGTCGGATTGGGAGAAAGAGGACTGCTGCTACTACACCATCTA 3411  
 QY 1103 TGCCGTTGAGGAGAGTGTGACCCCAATGCTGGTTTGAAGTCAAGAGGAGCTGGGAGA 1162  
 Db 3410 TGCAGTTGAAGCAGATGCTGACCAATCGAGCTTTGAAAAAACAAGAACCCAGGAGA 3351  
 QY 1163 AATACAGATATCTAATTAAGTGAAGAGCTGGTGCACACATCCATAACACTTGGGAACTGA 1222  
 Db 3350 GATTCAGATATTAATTAATTAAGTGAAGAGGATGGTCCCATATCCACACACTTGGGAGACAGA 3291  
 QY 1223 AGAACCTCTGAAGCAACAAATGTTAAAGGAATGACAAACTGGCACTACAAAGAAAA 1282  
 Db 3290 AGAACCTCTGAAGCAACAAATGTTAAAGGAATGACAAACTGGCACTACAAAGAAAA 3231  
 QY 1283 GGATCAGGACAAACCGCTGGCTGAAAAATGCTTCTCCAGAGAGTGTGGAAATATTATA 1342  
 Db 3230 AGATCAGGAACAAAGAGATGTTGAAAAATGCTTCTCCAGAGAGTGTGGAAATATTATA 3171  
 QY 1343 CTGCACGAGGAGCTTACAGATGATCTGCACAAACATATCAATAGTGGAAAGATTAAT 1402  
 Db 3170 TTGGCAGCAAGAACTTACAGATGATCTACATAACAGTATCAATAGTGGAAAGATTAAT 3111  
 QY 1403 TGCTCATTCAATCAAAAGTCAGCAGCTGGTTATCCGAGTACTATTTCAAAATGGCAGGG 1462  
 Db 3110 TGCTCATTCAATCAAAAGTCAGCAGCTGGTTATCCGATTTACTGCAATGGCAGGG 3051  
 QY 1463 TCTCCCTTACTCAGATGATGAGTGGGAGAGTGGTCTCTCATTTGCCAAGAGTTTCAGGC 1522  
 Db 3050 CTTTCCATCTCAGACTGAGTGGGAGAGTGGAGTCTCTCTCTTCCAAAGTTTCAGGC 2991  
 QY 1523 AGCAATGATGAGTATTTAGCAGAAATCAATCCAGACTACTCCCTTTAAAGGAGCTGCAA 1582  
 Db 2990 ATGCATTTGATGATATTTAGCAGAAATCAATCCAGACTACTCCCTTTAAAGGAGTTCGAA 2931  
 QY 1583 GGTCTTAAACAGAGACCAAGATTTGTCAGCTGAGAGCAAC 1626  
 Db 2930 ATGTAAGAAGAAGATGAGGCGCTGTTAAGCAGCTTTGAAACCAAC 2887

RESULT 6  
 AA280469/c  
 ID AA280469 standard; cDNA; 575 BP.  
 XX  
 AC  
 XX  
 XX  
 XX  
 07-APR-2000 (first entry)  
 DE Human colon cancer cell line SW480 cDNA clone SEQ ID NO:553.  
 XX  
 KW Human; gene expression product; diagnosis; tumour; colon cancer;  
 KW colorectal adenocarcinoma; cell line SW480; cell proliferation;  
 KW cytostatic; sarcoma; breast cancer; neoplasia; dysplasia;  
 KW hyperplasia; ds.  
 XX  
 XX Homo sapiens.  
 OS  
 PN W09964576-A2.

XX 16-DEC-1999.  
 PD  
 XX 09-JUN-1999; 99WO-IB01062.  
 PF  
 XX 10-JUN-1998; 98US-0088801.  
 PR  
 XX (FARB ) BAYER CORP.  
 PA  
 XX Endege WO, Steinmann KE, Astle JH, Burgess CC, Bushnell SE;  
 PI Carroll E, Catino TJ, Derti A, Ford DM, Lewis ME, Monahan JE;  
 PI Schlegel R;  
 DR WPI: 2000-087220/07.  
 XX  
 XX Novel nucleic acids, used to develop products for the diagnosis and  
 PT treatment of disorders involving unwanted cell proliferation,  
 PT particularly cancers, especially colon cancer  
 XX  
 XX Claim 15: Page 350; 469pp; English.  
 PS  
 XX AA279917 to AA280766 represent double stranded cDNA clones isolated from  
 CC the human colorectal adenocarcinoma (colon cancer) cell line SW480. The  
 CC cDNA clones can be used to generate antisense oligonucleotides which  
 CC can be used for antisense therapy. Methods and products from the present  
 CC invention can be used for identifying and/or classifying cancerous cells  
 CC present in a human tumour, particularly in solid tumours, e.g.  
 CC carcinomas and sarcomas, e.g. breast or colon cancers. The cDNA clones  
 CC can be used for developing agents for the diagnosis and treatment of  
 CC disorders involving unwanted cell proliferation, such as neoplasia,  
 CC dysplasia or hyperplasia.  
 XX  
 XX Sequence 575 BP; 187 A; 109 C; 95 G; 181 T; 3 other;

Query Match 4.9%; Score 321.4; DB 21; Length 575;  
 Best Local Similarity 85.4%; Pred. No. 3.3e-60;  
 Matches 358; Conservative 0; Mismatches 61; Indels 0; Gaps 0;  
 QY 2035 TCAATGGGATTCATGAGAGTGTGATGAGCTCATCGTTTAAATAATGATGACTCTCTTC 2094  
 Db 419 TAAATGGGATTCATGAGAGTGTGATGAGCTCATCGTTTAAATAATGATGACTCTCTTC 360  
 QY 2095 TGTACAGAGCTTTAATAGACTTTAAGTCCAAACCATGACCTTCGATGATGAGAGCTTCTC 2154  
 Db 359 TGTATAAACTTTAATAGACTTTAATGATGAGCTCATCGTTTAAATAATGATGACTCTCTTC 300  
 QY 2155 TGCATAAACTTTAATAGACTTTAAGTCCAAACCATGACCTTCGATGATGAGAGCTTCTC 2214  
 Db 299 TACAGATTTCCCTCAAGAGCTTGTGCTTTGCTACATTTTGTGCTACATTTTGTGCTACATTT 240  
 QY 2215 CCTCTGGGAAGATTTTGAAGAGGAGCATGGCAAGAGAGAGAGTATGTTTATGCAAGTC 2274  
 Db 239 CTTCCTGGGAAGATTTTGAAGAGGAGCATGGCAAGAGAGAGAGTATGTTTATGCAAGTC 180  
 QY 2275 TGCATAAACTTTAATAGACTTTAAGTCCAAACCATGACCTTCGATGATGAGAGCTTCTC 2334  
 Db 179 TGCATAAACTTTAATAGACTTTAAGTCCAAACCATGACCTTCGATGATGAGAGCTTCTC 120  
 QY 2335 TACCTGCTGAGGTTGAACAAATTTCTGAGGATGGAATGAGTGCATTTGCAAGAGCAATATT 2394  
 Db 119 TTCTCTGCCAAGTTGAGCAGAGATTTTAAAGATGGAATGAGTGCATTTTACAGAAACATATT 60  
 QY 2395 ACAAGTGTGATTTTAAACAGGAGTATTAAGCCCTCAGTAAAGGTTTCAAGAGCAGTACC 2453  
 Db 59 ACAATGATTTTAACTAGGATTTACAAAGCCCTCAGCAAGGTTTCCAGGCGAGTACC 1

RESULT 7  
 AA158444  
 ID AA158444 standard; cDNA; 6475 BP.  
 XX  
 AC AA158444;











Db 750 caacattgaaatgatcctagcagactgcccagagttgaaagaagattcaactggagctgtg 809  
QY 2050 TAGGAGTGTGATGAAGCTCATCGTTTAAATAATGATGACTCTCTCTGTCAGGACTTTAA 2109  
Db 810 tgataattgatgaagccacagactgaagaataggaaacttcttgagggtctaa 869  
QY 2110 TAGACTTTAAGTCCAACCACTGACTTCTGATTACTGGAACCCCACTGCAAAATTCCTCTCA 2169  
Db 870 agcttatggccctggaacataaagtgcttctcactggaacacccttgacgaactctgtg 929  
QY 2170 AAGAGCTGTGCTTGTGTTGCATTTTCATCATGCCAGAAAATTTTCTCTCTGGAGAGATT 2229  
Db 930 aggagctctcagttgttgaattttctggagccatcacagtttctctcagagaccgctt 989  
QY 2230 TTGAAGAGGAGCATGGCAAGAGAGAGATGTTATGCAAGT---CTTCACAAAGAGC 2286  
Db 990 tcttgagggaatttgagatctgaaacagagagcaggtaaagaactcgactctatcc 1049  
QY 2287 TTGAACCATTTTACTAAGAAGAGTGTAAATAAGATGTAGAAAAGTCTTTTACCTGCTAAGG 2346  
Db 1050 taaaccaatgatgcttcgcggtgaaagatgatgtggaaaagaaccttgcctcccaaac 1109  
QY 2347 TTGNACAAATTCGAGGATGGAATGAGTGCATTCGAGAGCAATATTACAAAGTGATTT 2406  
Db 1110 aagagacgatcatgaggtggaaactgacccaatatccagaaaagtactaccgtgccaacc 1169  
QY 2407 TAACAGGAATATTATAAGCCCTCAGTAAAGGTTCAA---AAGCGAGTACCTCAGGCTTTC 2463  
Db 1170 tcgagaagaacttctcctcctgacccaaggggcaaatcagcacaacatgcccacatctca 1229  
QY 2464 TGAACATTTATGATGGAACCTTAAAGTGTGTGAACCATGCTACCTCATTTAAGCCCAACC 2523  
Db 1230 tcaacacatgatggagctggaagtgctgtaaccatccctacctgacatcaatggtgcag 1289  
QY 2524 ATGATAATGAATCTATAATAACAGAGGCGCT----- 2556  
Db 1290 aggagaaaattcagaagattctccgaaaaacccacagccctgatgccctgactttcagc 1349  
QY 2557 TAGACGATTTGATACGTAGCAGCGGGAATACTAATCTCTTGACAAGCTACTGATTCGTC 2616  
Db 1350 tgcaggccatgattcagcagcaggaagctgtgttgattgataaactactccctaaagc 1409  
QY 2617 TGGAGAACCTGGCAACAGATTCGATTTCTCAGATGGTGGAGGATGCTGGACATCC 2676  
Db 1410 tgattgcagtggtgcccacaaagtactatcttcccagatgggtgcgtgcctgcacatcc 1469  
QY 2677 TAGCAGAAATATCTGAAGTATCGCAGTTTCCCTTCAGAGACTGTGATGGATCAATAAAG 2736  
Db 1470 tagaagatccctcctccagagaaggtacacccctatgagcgaaactgatggcgagctacgg 1529  
QY 2737 GGAATTTGAGGAAGCAAGCACTGGATCATTTCAATGCAGAAGGATCAGAGGATTTCTGTT 2796  
Db 1530 gaaacctgcgcaggcagccatgcacgggtgtctgtaagccagattcagaccgcttgtct 1589  
QY 2797 TTTTACTGCTTACAAGCTGGAGGATAGGATTTAATCTTTGATGATGCTGTCGACACTGTAG 2856  
Db 1590 ttctctgtgcaccagagcggggtggtggtatcctcctacatctcacagctgctgatacctgca 1649  
QY 2857 TTATTTTGTGATCTGACTGGAATCCACAGAAATGATCTGCAGGCGACAGCGAGCTCATA 2916  
Db 1650 tcatattgtattcgtggaacccaaatgacttgacggtcctcaggtccggtggtgaccc 1709  
QY 2917 GAATGAGACAGAAAGAGGTTTAATATTATCGGCTAGTACAAAAGGATCAGTAGAAG 2976  
Db- 1710 gcataggccagagcaaacgtgtgaaggtgtatcgctcctcactcgaattcctcagacg 1769  
QY 2977 AAGATATCTTGAAGAGCCCAAGAAAGATGCTGTAGACCAATTTAGTAATTTACAGAA 3036  
Db 1770 gcagatgtttgacagggccagcctaagctggggtggaagaggtgttcttctcaggaca 1829  
QY 3037 TGGAC 3041  
Db 1830 tcaac 1834

RESULT 11

ABL06515  
ID ABL06515 standard; cDNA; 8112 BP.

XX  
AC ABL06515;

XX  
DT 26-MAR-2002 (first entry)

XX  
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 14027.

XX  
KW Drosophila; developmental biology; cell signalling; insecticide;

XX  
KW pharmaceutical; gene; ss.

XX  
OS Drosophila melanogaster.

XX  
PN WO200171042-A2.

XX  
PD 27-SEP-2001.

XX  
PF 23-MAR-2001; 2001WO-US09231.

XX  
PR 23-MAR-2000; 2000US-191637P.

PR  
PR 11-JUL-2000; 2000US-0614150.

XX  
PA (PEKE ) PE CORP NY.

XX  
PI Venter JC, Adams M, Li PWD, Myers EW;

XX  
DR WPI; 2001-656860/75.

DR  
DR P-PSDB; ABB62412.

XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions -

XX  
PS Claim 1; SEQ ID NO 14027; 21pp + Sequence Listing; English.

XX  
CC The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA sequences (ABL01840-ABL16175) and the encoded proteins (AB5737-AB572072).

CC  
CC The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

CC  
XX Sequence 8112 BP; 2500 A; 2042 C; 2024 G; 1546 T; 0 other;

Query Match 4.0%; Score 266.8; DB 23; Length 8112;  
Best Local Similarity 52.6%; Pred. No. 7.5e-48;  
Matches 757; Conservative 0; Mismatches 627; Indels 54; Gaps 6;

QY 1660 TAGAGATTATCATGTTAAATGGATTGAATGGCTCGCTCATTCATGTCAGAGAAATA 1719  
Db 6065 tagaccctatcactgaggtcttaattggtgaagtttcttggtacacaccccaca 6124

QY 1720 GTTGTATTTCTGCAGATGAAATGGTCTGGTAAACAAATACAAACAAATTTCTTTCTGA 1779  
Db 6125 actgcatactcgtcagaaaatggccttggaataactattcaagtttgacatttgctc 6184

QY 1780 ACTTACTGTTTATGAAACATCAACTGTATGGCCCTTTCTTCTGCGCGTCCCACTTCTA 1839  
Db 6185 actctgtat---atgagtacggcattagaggaaccttcttagttatagctcctctatcta 6241

QY 1840 CCTTGACATCTTGCAAGAGAGATTCAAACTTGGGCTCCTCAGATCAATGCTGTAGTTT 1899  
Db 6242 caattccaaactggcagcgagagttcagggtggtg---acggatatgaacgtagttgtt 6298

QY	1900	ACTTAGGAGATAACTAGTAGAATAATATCATAAAGGACTCATCAATGGATGCATCC-----	1954
Db	6299	atcacggtcggtgcacagtaaacaaatgacagagactatgaatatctataaagacag	6358
QY	1955	-----ACAGACTAAACGATTAAAGTTTAAACATACCTCTGACGACATGAAA	2001
Db	6359	aaagtgaaggtattgaaggcccatcaagttaacgttttgatcaccacottttgaaa	6418
QY	2002	TTTTTACTGAAGGATAACTCATTCCTTGGTGGTCFCAATTTGGGCATTCTATAGGAGTTGATG	2061
Db	6419	tgaatgcagactacaatggacttaaaagcctttaacggcgccttgcgtgatgcag	6478
QY	2062	AGCTCATCGTTTAAAAATATGACTCTCTCTGTCAGAGACTTTAATAGACTTTAAGT	2121
Db	6479	aggcacatcgctttaagaaatgaagattgcaaacctccttgagggtctgcacagttaaatt	6538
QY	2122	CCAAACATCGACTTCTGATTACTTGAACCCACCTGCMAAATTTCCCTCAAGAGACTGTGGT	2181
Db	6539	tgagcacagagtatgtctctcggaactccctcaaacacaacatcagcgagctgtct	6598
QY	2182	CTTTTGTGCAATTTCATGCCAGAAAAATTTCTCTCTGGGAAGATTTTGAAGAGGAGC	2241
Db	6599	cgctgttaacttctggaacctcgcagttctctccaggaagatcctatctgagt	6658
QY	2242	ATGCCAAGGAGAGATGTGTTATGCAAGT---CTTCACAAGAGCTGTGAACCAATTTT	2298
Db	6659	tgggaagcttcgcactgagagaagaatataatagctgcgaagctctactgaaacaaatga	6718
QY	2299	TACTAGAGAGATTAAAAAGATGTAGAAAAAGTCTTTACTCTGAAGTTGAAACAAATTC	2358
Db	6719	tgatcgtcgtctaaagacagctagagaaagtgtggcgcccaaggaagaaacattat	6778
QY	2359	TCAGGATGGAATAGTGCATTTCGCAAGCAATATTACAAGTGGATTTTAACAAAGAAATT	2418
Db	6779	tcgaagtggagctaacataacatacaagaataattatcgaggtatactgagaaacgaact	6838
QY	2419	ATAAAGCCCTCAGTAAGAGTTCAAAA---GGCAGTACCTCAGGCTTTCTCAACATTATGA	2475
Db	6839	ttagttctctgaaaaagggaccacatctgtaatatcccaaaccttatgaacacatga	6898
QY	2476	TGGAACTTAAGAAGTGTGTAACCATTTGCTACTCTCATTAAGCCACGAGTGAATGAAT	2535
Db	6899	tggaattgagaagtgcgtacacaccttatctcttgatggagcgagggaacaaact	6958
QY	2536	TCATAATTAACAGCA-----GGCCTTACGATTTTGATTC	2571
Db	6959	aatatgatttccaagtcccaacgaatgcgcgaagatcctgagtcctattataaaatctgattc	7018
QY	2572	GTAGCAGCGGGAACATAACTCTTTCTTGACAAGCTACTGATTTCGTCTCGAAGACGTGCA	2631
Db	7019	tctccgctggtaaaaatggttttaattgataaattgctaactaaacataaagcaaacgcgc	7078
QY	2632	ACAGAGTTTCTGATTTTCTCTCAGATGGTGAGGATGTGGACATCCTAGCAGAAATCTGA	2691
Db	7079	atcgctctctaatactcagtcagatggcgctgtgcttgatatactctgagattatctag	7138
QY	2692	AGTATGCCAGATTTCCTTCACAGACTTGATGGATCAATAAAGGGGAATTCAGGAAGC	2751
Db	7139	tgtacagaaaaatcacccctttgagcgaatcgatgggcgcattcgcgcgttaactctccgcag	7198
QY	2752	AAGCACTGGATCAITTCANTGTCAGAGAGNATCAGAGATTTCTGTTTTTACTGCTACAA	2811
Db	7199	agggcaatgatcgtttactcccaacgaaggtccgactgcgtttgattctcctcttgacaa	7258
QY	2812	GAGCTGGAGATTAGGTATTAACCTTGGCATCTGCTGCACACTGTAGTTATTTTGAFTCTG	2871
Db	7259	agcgtggtggaatagttatlaatttaacagctgctgatactgttatattatcagatctcg	7318
QY	2872	ACTGGAATCCACAGAATGATCTCGAGGCACAGCGGAGAGCTCATAGAATTTGACAGAGA	2931
Db	7319	attggaattcacaaaqaatttacagactcagqccgcagtcacatcgatttgccgcagaa	7378

Qy	2932	ACAGGTTATATTTATTCGCTAGTCACAAAGGATCAGTAGAAGATATATCTTGAA	2991
Db	7379	agaTggtaagaatttatagattgcttttcaggaaatacctatgagcgtgaattgttcaca	7438
Qy	2992	GAGCCAAAGAAGAATGGTGCTAGACCAATTTAGTAATTCACGAAGATCGACACGAGG	3049
Db	7439	agcgttcaatgaaacttgcattgacaaagcgtctattgcattcatcaacacccaaq	7496

RESULT 12

ABL06514	
ID	ABL06514 standard; cDNA; 10274 BP.
XX	AC
XX	AC
XX	ABL06514;
XX	
DT	26-MAR-2002 (first entry)
XX	
DE	Drosophila melanogaster expressed polynucleotide SEQ ID NO 14024.
XX	
XX	Drosophila; developmental biology; cell signalling; insecticide;
KW	pharmaceutical; gene; ss.
KW	
XX	Drosophila melanogaster.
OS	
XX	WO200171042-A2.
FN	
PD	27-SEP-2001.
XX	
XX	23-MAR-2001; 2001WO-US09231.
XX	
XX	23-MAR-2000; 2000US-191637P.
PR	
PR	11-JUL-2000; 2000US-0614150.
XX	
XX	(PEKE ) PE CORP NY.
PA	
XX	
PI	Venter JC, Adams M, Li PWD, Myers EW;
PI	
PI	WPI: 2001-656860/75.
DR	P-PSDB; ABB62411.
DR	
XX	
PT	New isolated nucleic acid detection reagent for detecting 1000 or more
PT	genes from Drosophila and for elucidating cell signalling and cell-cell
PT	interactions -
XX	
XX	Claim 1; SEQ ID NO 14024; 2ipp + sequence Listing; English.
PS	
XX	
XX	The invention relates to an isolated nucleic acid detection reagent
XX	capable of detecting 1000 or more genes from Drosophila. The invention
CC	useful in developmental biology and in elucidating cell signalling and
CC	cell-cell interactions in higher eukaryotes for the development of
CC	insecticides, therapeutics and pharmaceutical drugs. The invention
CC	discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC	sequences (ABL01840-ABL16175) and the encoded proteins
CC	(ABB57737-ABB72072).
CC	The sequence data for this patent did not form part of the printed
CC	specification, but was obtained in electronic format directly from WIPO
CC	at ftp.wipo.int/pub/published_pct_sequences.
XX	
XX	Sequence 10274 BP; 3117 A; 2515 C; 2418 G; 2224 T; 0 other;
XX	

Query Match 4.0%; Score 266.8; DB 23; Length 10274;  
Best Local Similarity 52.6%; Pred. NO. 8.2e-48;  
Matches 757; Conservative 0; Mismatches 627; Indels 54; Gaps 6;

Qy	1660	TAAGAGATTATCAGTTAAATGGATGAAATGGCTGCGCTCATTCATGGTGCAAAAGGAATA	1719
Db	7165	ttgaccctatcaactgagggtcttaatgtttgaagtttcttggacaacaccacca	7224
Qy	1720	GTTGTATTCTTCGAGATCAAAATGGGTCTGGGTAAAAACAATCAACAATTTCTTTCTTGA	1779
Db	7225	actcacaactcactcaccaaatgggccttggaaaaactatcaaaqtttgacattttgctg	7284

Qy	1780	ACTTACCTGTTTCTATGAAACATCAACTGTATGCGCCCTTTTCTTCTCTGCGCTGCCACTTCTTA	1839
Db	7285	actctgcat---atgagtcgcgcattagagagaccttctctagttatagctcctctatata	7341
Qy	1840	CCTTGACATCTTCGCAAGAGAGAGATTCAAACTTGGGCTCCCTCAGATGAATCCTAGTGT	1899
Db	7342	caatccaaactggcgagaggttcgagggctgg---acggatatgaocgtagttgttt	7398
Qy	1900	ACTTTAGGAGATATAAATCTAGTAATAATATCAATGAGACTCATCAATGGATGTCATCC	1954
Db	7399	atcaaggctcggfagcaagtaaacaaatgatcacagagactatgaatatctactataagacag	7458
Qy	1955	-----ACAGACTAAACGATTTAAAGTTTAACTACTTCTTGACGACATATGAAA	2001
Db	7459	aaagtggaaaggfatiagaaggagcccatcaagtttaaagcttttgatcacacctttgaaa	7518
Qy	2002	TTTTTACTGAAGGATAAGTCATTCCTTGGTGGTCTCAATTTGGGCATTCATAGAGGTTGATG	2061
Db	7519	tgaicgagacactacatggacttaaaagcccttcaactggcgcccttctgttggattgatg	7578
Qy	2062	AAGCTCATCTGTTTAAAAAATGATGACTCTCTCTGTACAGGACTTTAATAGACTTTAAGT	2121
Db	7579	aggcacatcgtttaagaataggaattggcaactccttggaggctcgacagttaaatt	7638
Qy	2122	CCAAACCATGACATTCGATTACTTGGNAACCCCACTGCAAAATTCCTCAAGAGCTGTGGT	2181
Db	7639	tggagcacagatgatctctccgggaactcccctacaaaacaacatcagcgagctgtct	7698
Qy	2182	CTTTTGTGCAATTCATCATGCCAGAAAATTTTCTCTCGGAGAGATTTTCAAGAGGAGC	2241
Db	7699	cgcgtttaactcttcggaaacctcgcagttctctccacaggaaggttcatgtctgagt	7758
Qy	2242	ATGGCAAAAGGAAGAGATATGTTTATGCAAGT---CTTCAACAAGAGCGTTCAACCAATTT	2298
Db	7759	ttggaagtcttcgcactggaagaagaagtaataaagtgcgaagctctactgaacccaatga	7818
Qy	2299	TACTAAGAAGAGTTTAAAAAGATGTAGAAAAGCTTTTACCTGCTAAAGTTGGAACAAATTC	2358
Db	7819	tgttaactgctctaaaagacgacgtagaagaagtttggcgcccaaggagaagaacccaatta	7878
Qy	2359	TGAGGATGGAAATGACTGCATTCCGAGACCAATATTACAAGTGGATTTTACACAGGAATT	2418
Db	7879	tcgaagtggaagtcacctaacaacaaaagaataatacgaaggtatatactgggaacagaact	7938
Qy	2419	ATAAAGCCCTTCAGTAAAGGTTCAAAA---GGCAGTACCTCAGGGTCTCTGAACATTTATGA	2475
Db	7939	ttagtttccctgaaaaggagggaccacatctgctaatatcccaaaccttatgaacacatga	7998
Qy	2476	TGGAACTTAAGAAGTGTGTTAACCATTTGCTACCTCATTAAGCCACCACAGATGAATGAAT	2535
Db	7999	tggagttgagaagtgctgcatacacaccttatctcttgaaatggagcgaggagaacaaatcc	8058
Qy	2536	TCATATATAACAGGA-----GGCCTTACAGCATTTTGATAC	2571
Db	8059	aatatgatattcaagtcccagcatggcgaaagatccttgagtccttatataaaatctgcttc	8118
Qy	2572	GTAGCAGCGGGAACATAACTCTTCTGTGACAAGCTACTGATTCTGCTCGGAGAAGTCGGCA	2631
Db	8119	tttccctcgtgaaaaatgggttttaattgataaattgctacctaaactaaaagcaaacggccc	8178
Qy	2632	ACAGAGTTCGATTTTCTCTCAGATGGTGAGGATGCTGGACATCCTTAGCAGAATATCTCGA	2691
Db	8179	atcgcgtcttaatttcagtcagatggcggttgcgttgcgttgatatccttgaagattatctag	8238
Qy	2692	AGTATCCCGAGTTTCCCTTCCAGAGACTTGATGGATCAATAAAGGGGGAATTTGAGGAAGC	2751
Db	8239	tgtacagaaaaataccccttctgagcgaatcogatggcgccgattcgcgglaactctccgcag	8298
Qy	2752	AAGCACTGGATCATTTCAATGCAAGAGGATACAGAGATTCTGTGTTTTTACTGCTACTCAA	2811
Db	8299	agggccaicgagcttactctccagcgaggtccogaicgcttctgtattctctctctgcacca	8358
Qy	2812	GAGCTGGAGGATTTAGGTATTTAACTTTGGCATCTGCTGACACATGTAGTATATTTTGTATCTG	2871

Db	8359	aagctggtgattaggattatttaatttaaacagctgctgatactgttattattacagattcgg	8418	
Qy	2872	ACTGGAATCCACAGAATGATCTGCAGGCACAGCGAGAGCTCATAGAAATTGGACAGAAGA	2931	
Db	8419	attggaattccacaaaacgatttacaggctcagggcccgatgccatcgatcggccagagaa	8478	
Qy	2932	AACAGGTTAATATTTTCGCGTAGTCACAAAAGGATCATAGTAGAAGAAGATATCTTGAAA	2991	
Db	8479	agatggtaagattttatagattgctttgcagggaataacctatgagcgtgaaatgttcgaca	8538	
Qy	2992	GAGCCAGAGAGAGATGGTGCTAGACCATTTAGTAATTCAGAGATGGACACGACAGC	3049	
Db	8539	aagcttcaatgaaacttgatggacaaggctgtattgcagtcgatgaacacccaagg	8596	
RESULT	13			
ID	ABL14275			
XX	ABL14275 standard; cDNA; 3700 BP.			
AC	ABL14275;			
DT	26-MAR-2002 (first entry)			
XX	Drosophila melanogaster expressed polynucleotide SEQ ID NO 37307.			
DE	Drosophila; developmental biology; cell signalling; insecticide;			
KW	pharmaceutical; gene; ss.			
XX	Drosophila melanogaster.			
OS	WO200171042-A2.			
PN	27-SEP-2001.			
PD				
XX	23-MAR-2001; 2001WO-US09231.			
PF				
XX	23-MAR-2000; 2000US-191637P.			
XX	11-JUL-2000; 2000US-0614150.			
PR	(PEKE ) PE CORP NY.			
XX	Venter JC, Adams M, Li PWD, Myers EW;			
FI	WPI; 2001-656860/75.			
XX	P-PSDB; ABB70172.			
DR	New isolated nucleic acid detection reagent for detecting 1000 or more			
PT	genes from Drosophila and for elucidating cell signalling and cell-cell			
PT	interactions -			
XX	Claim 1; SEQ ID NO 37307; 21pp + Sequence Listing; English.			
PS	The invention relates to an isolated nucleic acid detection reagent			
XX	capable of detecting 1000 or more genes from Drosophila. The invention			
CC	useful in developmental biology and in elucidating cell signalling and			
CC	cell-cell interactions in higher eukaryotes for the development of			
CC	insecticides, therapeutics and pharmaceutical drugs. The invention			
CC	discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA			
CC	sequences (ABL01840-ABL16175) and the encoded proteins			
CC	(ABB57737-ABB72072).			
CC	The sequence data for this patent did not form part of the printed			
CC	specification, but was obtained in electronic format directly from WPI			
CC	at ftp.wipo.int/pub/published_pct_sequences.			
XX	Sequence 3700 BP; 1052 A; 861 C; 940 G; 847 T; 0 other;			
SQ				

Query Match 3.7%; Score 245.8; DB 23; Length 3700;  
Best Local Similarity 50.5%; Pred. No. 2e-43;  
Matches 706; Conservative 0; Mismatches 657; Indels 36; Gaps 3



CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
CC in gene therapy. A composition containing a polypeptide or polynucleotide  
CC of the invention may be used to treat diseases of the peripheral nervous  
CC system, such as peripheral nervous injuries, peripheral neuropathy and  
CC localised neuropathies and central nervous system diseases, such as  
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
CC utilisation of the activities such as: Immune system suppression,  
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
CC assays for receptor activity, arthritis and inflammation, leukaemias and  
CC C.N.S disorders.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification.

SQ Sequence 6553 BP; 1802 A; 1538 C; 1829 G; 1384 T; 0 other;

Query Match 3.6%; Score 239.4; DB 22; Length 6553;  
Best Local Similarity 52.1%; Pred. No. 6.3e-42;  
Matches 744; Conservative 0; Mismatches 621; Indels 62; Gaps 7;

QY	1568	TATCAGTTAAATGGATGTAATTTGGCTCGCTCATTCATCGTGCAAGGAATAGTTGTATT	1727
Db	2383	tatacaatggaggccgaattggttgcgctctcctggtcgctcaggcgcaatgacacattc	2442
QY	1728	CTTGCAGATGAATGGTCTGGGTAAACAATACAAAACAATTTCTTTCTGAACTACCTG	1787
Db	2443	ttggcctgctgagatgggccccttggaaaaactgtacagacagcagctctctcgttatccctt	2502
QY	1788	TTTTCATGAACATCAACTGTATGGCCCTTTCTTCCTGCGCGTGCACATTTCTTACCTTCA-C	1846
Db	2503	tacaaggaggctcattccaaaagcccctctctagtgcgagccctctttctaccatcatc	2562
QY	1847	ATCTTGGCAAAAGAGAGATTCAAACTTTGGCTCCTCAGATGAATG-CTGTAGTTTACTTTAG	1905
Db	2563	aacttgaggcggaatttgaattgggctccagacatgatgtcogtaacctatgtg	2622
QY	1906	GAGATTAACCTAGTACAAATATGATAAGGACATCATGAATGNTGCATCCACAGACTAAAC	1965
Db	2623	gtgacaagagacagccgtgcattccagagagatgagttctccctttgaaagacaatgccc	2682
QY	1966	GAT-----TAAAGTTTAAACATA	1982
Db	2683	attcgtgtgggcaagaagccctccgcattgaagaagaaggcatctgtgaaattccattgtg	2742
QY	1983	CTTCTGACGACATATGAAATTTTACTGAAAGATAAGTCAATTCCTTGGTGGGTCAATTTGG	2042
Db	2743	ctgctgacatccctaatgtatgtaccattgacatggctattttggcgctctattgattgg	2802
QY	2043	GCATTCATAGGAGTTCATCAAGCTCATCGTTTAAAAAATGATGACTCTCTCTGTACAGG	2102
Db	2803	gcccgtctcatcgtggatgaagcccatcggctggaagaacaatcagttcctaagttcttcgg	2862
QY	2103	ACTTTAATAGACTTTAAGTCCAACCATCGACTCTCTGATTACTGAAACCCCACTGCAAAAT	2162
Db	2863	gtattgaaagttaactcactccagcacaagctgttgctgcactggagacccaattacaaaac	2922
QY	2163	TCCTCTAAGAGCTGTGGTCTTTTGTGCAATTCATCTGCCAGAAAAAATTTTCTCTCTGG	2222
Db	2923	aatctggaagagttttctcatctgtctcaacttctcaccgccgagaggttccacaattg	2982
QY	2223	GAAGATTTGAAGAGGAGCATGCAAGGAAGAGAGATGCTTATGCAAGTCTTCACAAA	2282
Db	2983	gaggggttttttggaggagtttctgctgacattgccaaaggaggaccagataaaaaaactgcat	3042
QY	2283	GAGCT----TGAACCAATTTTACTAAGAGAGATTAAAAAAGATGTAGAAAAAGTCTTTACC	2338
Db	3043	gacatgctggggccgcacatgtgtcggcggtcctcaagccgcatgtgttccaagaacatgccc	3102
QY	2339	TGCTAGGTTGAACAAATTTCTGAGGATGGAATAGTGC--ATTGCGACAGCAATATATAC	2396
Db	3103	ctccagacagacaactaattgtgtctgtggagctgagccctatgacagaagaatactac	3162

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX WPI: 2001-318749/34.  
XX  
XX Primer sets for synthesizing polynucleotides, particularly the 5602  
PT full-length cDNAs defined in the specification, and for the detection  
PT and/or diagnosis of the abnormality of the proteins encoded by the  
PT full-length cDNAs -  
XX  
XX Claim 8; SEQ ID 14295; 2537pp + CD ROM; English.  
XX  
XX The present invention describes primer sets for synthesizing 5602  
CC full-length cDNAs defined in the specification. Where a primer set  
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
CC to the complementary strand of a polynucleotide which comprises one of  
CC the 5602 nucleotide sequences defined in the specification, where the  
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
CC of an oligonucleotide comprising a sequence complementary to the  
CC complementary strand of a polynucleotide which comprises a 5'-end  
CC sequence and an oligonucleotide comprising a sequence complementary to a  
CC polynucleotide which comprises at least 15 nucleotides and the combination of  
CC the 5'-end sequence/3'-end sequence is selected from those defined in  
CC the specification. The primer sets can be used in antisense therapy and  
CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to  
CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632  
CC represent oligonucleotides, all of which are used in the exemplification  
CC of the present invention.  
XX  
XX Sequence 2195 BP; 718 A; 391 C; 494 G; 592 T; 0 other;

Query Match 3.6%; Score 236.6; DB 22; Length 2195;  
Best Local Similarity 51.9%; Pred. No. 1.7e-41;  
Matches 713; Conservative 0; Mismatches 609; Indels 51; Gaps 6;

Db 418 tgactggcaccctctccaaatacacagtgaagaactatttagtcttctcactttctg 477  
QY 2200 GCCCAGAAAATTTTCTCTGGGAGAGATTTTAAAGAGGAGCATGGCAAGAGAGAGT 2259  
Db 478 aaccttaaggtttctctctgaatcaacattttgtcaagaatttggggctctgaaacag 537  
QY 2260 ATGGTTATGCA--AGTCTTCACAAAGAGCTTGAACCATTTTAAAGAGAGTAAAA 2316  
Db 538 aggaacagggtacagaacttcaggttatcttgaaacccaatgatgtgagagattaaag 597  
QY 2317 AGATGTAGAAAGTCTTTACCTGCTAAGTGTGAACAAATCTGAGAGTGAATAGTG 2376  
Db 598 agatgtggaagaagaagttggccacctaaagaagaacacattgaagtgaactacta 657  
QY 2377 CATTCGAGAGAGCAATATTACAGTGGATTTTAAAGGAATTTAAAGCCCTAGTAAAG 2436  
Db 658 atattcaaaagaatactaccgggtctatcttggaagaagaactttcttttttaccaaag 717  
QY 2437 GTTCAAAGCGAGTACCTCA---GGCTTTCTGAACATTTATGATGAAGCTTAAAGTGT 2493  
Db 718 gagcagacaaactaatgtaccttaacttggctcaataccatgatggagctcagggaatgtt 777  
QY 2494 GTAAACATTGCTACCTCAATTAAGCCACAGATGATATGAATTCATATAAACAGAGG 2553  
Db 778 gtaatcatccatcttataaagggtgctgaggagaaataacttggagaattagagata 837  
QY 2554 CCT-----TACAGCATTTGATACGTAGCAGCGGAAAC 2586  
Db 838 cttaacaatcagctgtctctgatttctcatcttcaagcaaatgaccagctgtctggttaaat 897  
QY 2587 TAATCCCTTTTGACAGAGCTACTGATTCGCTGCGAGACGTGGCAACAGAGCTGTGATTT 2646  
Db 898 tggctctattgataaattgtctcccaaaatgaagcggaggtcataaagtgctcatct 957  
QY 2647 TCTCTCAGATGGTGGAGTCTGCAGATATCTAGCAGATATCTGAAGTATGCCAGTTTC 2706  
Db 958 tctctcaaatggtgtgcttgaccttgagagactctctacataaagaatact 1017  
QY 2707 CTTTCGAGAGACTTGATGATCAATAAAGGGGAATTTGAGGAAGCAAGCACTGATCATTT 2766  
Db 1018 tatatggcgaattgatggcagagctcagaggaatactctcggcaagctctcagatagat 1077  
QY 2767 TCAATGAGAGAGTACAGAGATTTCTGTTTTTACTCTCTACAGAGCTGGAGGATTAG 2826  
Db 1078 ttatgaacctgtatccagatagattgttttctctctgtgtaccagagctgggtgtg 1137  
QY 2827 GTATTAATTTGGCATCTCTGACACTGTAGTATTATTTTGTGATCTGGAATCCACAGA 2886  
Db 1138 gcatcaacttaactgcagctgatacatgtataattttgtctgtggaatcctcaga 1197  
QY 2887 ATGATCTGACAGGACAGGAGAGCTCATAGAAATTTGGACAGAAACAGAGTAAATATT 2946  
Db 1198 atgatcttcaggcccaagctgtggccacagaattggtcagaagaacagcagttaaagtct 1257  
QY 2947 ATCGGCTAGTCACAAAGGATTCAGTAGAAGAAGATATTCTTTGAAGAGCCGAAGAAGA 3006  
Db 1258 acagactggtaactcgttaactcatalgagagagagatgtttgaccgagcagttgaaac 1317  
QY 3007 TGGTCTAGACCATTTTACTTAATTCAGAGAAATGGACACAGCAAGAAAACCTGTT 3059  
Db 1318 tgggcctagataaagctgtgttacagagagcatgagtggaagagagaataatgtt 1370

Search completed: August 2, 2002, 20:52:39  
Job time: 26250 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 2, 2002, 16:52:15 ; Search time 8940.62 Seconds  
(without alignments)  
9975.581 Million cell updates/sec

Title: US-08-973-363-10  
Perfect score: 6608  
Sequence: 1 CGGCGTGGCGACGACGCGC.....AACTAAACCAAAAAAAC 6608

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estim:\*  
5: em\_estov:\*  
6: em\_estro:\*  
7: em\_estro:\*  
8: em\_estro:\*  
9: gb\_estl:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_gss:\*  
13: em\_gss\_hum:\*  
14: em\_gss\_inv:\*  
15: em\_gss\_pin:\*  
16: em\_gss\_vrt:\*

ALIGNMENTS

RESULT 1  
AKO15218  
LOCUS  
DEFINITION  
Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:4930428D05:chromodomain helicase DNA binding protein 1, full insert sequence.  
1811 bp mRNA linear HTC 19-JAN-2002

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
AKO15218  
AKO15218.1 GI:12853470  
HTC; CAP trapper.  
Mus musculus (strain:C57BL/6J) adult male testis cDNA to mRNA, clone.lib:RIKEN full-length enriched mouse cDNA library  
clone:4930428D05.

ORGANISM  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS  
Carninci, P., Hayashizaki, Y., Kondo, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y. High-efficiency full-length cDNA cloning Meth. Enzymol. 303, 19-44 (1999)

TITLE  
JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS  
Carninci, P., Hayashizaki, Y., Kondo, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes Genome Res. 10 (10), 1617-1630 (2000)

TITLE  
JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS  
Carninci, P., Hayashizaki, Y., Kondo, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes Genome Res. 10 (11), 1757-1771 (2000)

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Query Match %	Score	Length	DB	ID	Description
1	1144.4	17.3	1811	11	AKO15218	AKO15218 Mus muscu
2	629.6	9.5	648	10	BM491730	BM491730 pqp2n.pk0
3	581	8.8	888	10	BI657261	BI657261 603282944
4	575	8.7	597	10	BM486590	BM486590 pqp2n.pk0
5	473.6	7.2	1042	10	BM475844	BM475844 AGENCOURT
6	458.4	6.9	899	9	AU117528	AU117528 AU117528
7	456.6	6.9	866	9	AU125712	AU125712 AU125712
8	452.8	6.8	686	9	AW997058	AW997058 OV3-BN004
9	449.2	6.8	609	9	AW467566	AW467566 he21e07.x
10	440.4	6.7	895	10	BC538825	BC538825 602568268
11	432.8	6.5	1346	11	AKO18451	AKO18451 Mus muscu
12	432	6.5	1100	10	BE612384	BE612384 601451944
13	431.6	6.5	804	9	AV729238	AV729238 AV729238
14	431.4	6.5	808	9	AI326723	AI326723 mq55h09.y
15	422	6.4	765	10	BC485663	BC485663 802505962
16	412.6	6.2	633	10	BE427518	BE427518 df89cl1.y
17	412.6	6.2	753	10	BG167010	BG167010 602344974

[illegible]



DEFINITION 603282944F1 NTH\_CGAP\_Mam4 Mus musculus cDNA clone IMAGE:5327267 5', mRNA sequence.

ACCESSION BI657261

VERSION BI657261.1 GI:15571497

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 888)

AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

TITLE NIH-MGC <http://mgs.nhl.nih.gov/>.

JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)

COMMENT Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: [cgaps-remail.nih.gov](mailto:cgaps-remail.nih.gov)

Tissue Procurement: Lothar Hennighausen Ph.D., Priscilla Furth

Ph.D. Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Plate: LLML1830 row: g column: 12

High quality sequence stop: 741.

FEATURES

source

1..888

Location/Qualifiers

/organism="Mus musculus"

/db\_xref="taxon:10090"

/clone="IMAGE:5327267"

/clone\_lib="NIH\_CGAP\_Mam4"

/tissue\_type="tumor, gross tissue"

/lab\_host="DH10B"

/note="Organ: mammary; Vector: pCMV-SPORT6; Site:1: NotI; Site:2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert 2.5 kb. Library constructed by Life Technologies, catalog # 12018-016. Investigators providing samples: Lothar Hennighausen/Priscilla Furth, NIH Reference for transgenic model: Li et al., Cell Growth and Differentiation 7, 3-11 (1996). Note: this is a NCI\_CGAP Library."

BASE COUNT 270 a 185 c 203 g 230 t

ORIGIN

Query Match 8.8%; Score 581; DB 10; Length 888;

Best Local Similarity 82.8%; Pred. No. 1.9e-70;

Matches 688; Conservative 0; Mismatches 140; Indels 3; Gaps 2;

QY 2020 CATTCTTGGTGTCTCAATTGGGCATTCATAGGAGTTGATGAAGCTCATCGTTTAAAAA 2079

Db 2 CATTCTTGGTGTCTCAATTGGGCATTCATAGTGTGATGAAGCGCATCGATTAAAGA 61

QY 2080 ATGATGACTCTCTCTACAGACACTTAATAGACTTTAAGTCAACCATCGACTTCTGA 2139

Db 62 ATGATGATTCCTCTGTACAAAACCTTAATCGACTTTAAATCAACCATCGCTCTGA 121

QY 2140 TTACTGAACCCCACTCAAAATTCCTCAAGAGCTGTGCTTTGTCGATTTTCATCA 2199

Db 122 TCACGTGAACCCCTCTACAGACTCCCTGAAGAGCTCTGGTCACTCTGCATCTTATTA 181

QY 2200 TGCCAGAAAAATTTCTCTCTGGGAAGATTTTGAAGAGGAGCATGGCAAGGAGAGGT 2259

Db 182 TGCCGGGAAGTTTCTTCATCGGAAGATTTTGAAGAAGAACAATGGCAAGGAGAGAT 241

QY 2260 ATGTTTATGAAGCTTTCACAAAGAGCTTGAACCATTTTACTAAGAAGAGTTAAAAAG 2319

Db 242 ATGTTTATGAAGCTTTCACAAAGAGCTTGAAGCATTTCTGTACCGAGAGTTAAGAAG 301

QY 2320 ATGTAGAAAAAGTCTTTACCTGCTAAGGTTGAACAAATTTCTGAGATGCAATGAGTGCAT 2379

Db 302 ATGTGAAAAATCTCTCTCCGCAAGGTGGAGCAGATTTTAAGAAATGGAGATGAGTCTT 361

QY 2380 TGCAGAAGCAATATTACAACTGGGATTTTAAAGAAATTTATAAGCCCTCAGTAAGGTT 2439

Db 362 TACAGAAGCAATATTACAACTGGGATTTTAAAGAAATTTATAAGCCCTCAGCAAGGTT 421

QY 2440 CAAAGGCGAGTCTCAGGCTTTCTGAACATTTATGATGAAGCTTAAGAAGTTGTGAACC 2499

Db 422 CCAAGGCGAGTCTCAGGCTTTTGAACATTTATGATGAAGCTTAAGAAGTTGTGAACC 481

QY 2500 ATTGCTACCTCATTAAAGCCACAGATGATAATGAATTTCTATAATAACAGGAGCCCTTAC 2559

Db 482 ATTGCTACCTCATTAAAGCCACAGATGATAATGAATTTCTATAATAACAGGAGCCCTTAC 541

QY 2560 AGCATTGTATACGACGCGGAAACTAATCTCTTGTGAAAGCTACTGATGCTGCTGC 2619

Db 542 AACACTTAATCCCTAGTAGCGGAAACTCATCTCTCGACAAAGCTTTGATTCGCTAA 601

QY 2620 GAG-AACGTGGCAACAGAGTTCTGATTTCTCTCAGATGGTGAGGATGCTGGACATCTTA 2678

Db 602 GAGAAACGAGGCAACCGAGTGCTCATTTCTCTCAGATGGTGCGGATGTTAGACATCTC 661

QY 2679 GCAGAATATCTGAAGTATCGCCAGTTTCCCTTCCAGAGCTTGTGATCAATAAAAAAGG 2738

Db 662 GCAGAGTATTGAAGTACCCCGAGTCCCTTTCAAGATTAGATGATCGATAAAGGG 721

QY 2739 GAATTGAGGAAGCAAGCACTGGATCATTTCAATGCAGAAGATCAGAGGATTTCTGTTT 2798

Db 722 GAGCTGAGGAAGCAGGCTTGGATCCACTTTACTGCCGAGGCTCAGAGGATT--CTGCTT 779

QY 2799 TTACTGCTACAGAGCTGGAGGATTAGGTATTAACTTGGCATCTGCTGAC 2849

Db 780 ATGCTTTTAAACGAGCTGGGGGTTTCAGGCATAACCTAGTCTGTGAACAC 830

RESULT 4

BM486590

LOCUS

DEFINITION

BM486590 597 bp mRNA linear EST 07-FEB-2002

pgm2n.pk002.b16 Normalized Chicken Breast Muscle, Leg Muscle, and Epiphyseal Growth Plate cDNA library (pgm2n) Gallus gallus cDNA clone pgm2n.pk002.b16 5' similar to gb|AAC60282.1 (AF004397) chromo-helicase-DNA-binding on the 2 chromosome protein [Gallus gallus], mRNA sequence.

ACCESSION BM486590

VERSION BM486590.1 GI:18607520

KEYWORDS EST.

SOURCE chicken.

ORGANISM Gallus gallus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.

REFERENCE 1 (bases 1 to 597)

AUTHORS Cogburn,L.A. and Monsonogo-Ornan,E.

TITLE Epiphyseal Growth Plate cDNA library, USDA/IFAFS Animal Genome Project

JOURNAL Unpublished (2002)

COMMENT Contact: Larry A. Cogburn

University of Delaware

Townsend Hall, Newark, DE 19717, USA

Tel: 302-831-1335

Fax: 302-831-2822

Email: [cogburn@udel.edu](mailto:cogburn@udel.edu), [www.chickest.udel.edu](http://www.chickest.udel.edu).

FEATURES

source

1..597

/organism="Gallus gallus"

/strain="Commercial broiler and Ottawa Res. Centre Strains 90 & 21"

/db\_xref="taxon:9031"

/clone="pgm2n.pk002.b16"

/clone\_lib="Normalized Chicken Breast Muscle, Leg Muscle, and Epiphyseal Growth Plate cDNA library (pgm2n)"

/sex="Male and Female"

/tissue\_type="Breast muscle, leg muscle and epiphyseal growth plate"

```
/dev_stage="Breast,leg:Embryo(d19);post-hatch(1d,1,3,5,7,9,11 weeks);growth plate(1d,7d,14d post-hatch)"
```

Db 600 AACCATAGAAAGATTATGGATTGCGGATTGGGAGAAAAGAGCTACTGGTGTACTA 659  
QY 1093 CCACCATCTATGCGTGTAGCGACAGATGGTGACCAAAATGCTGGTGTGAAAAGTCAAAGG 1152  
Db 660 CAACCATCTATGCGTGTAGCGACAGATGGTGACCAAAATGCGAGCTTTGAAAAACAAGG 719  
QY 1153 AGCTGGGAGAAATACAGTATCTTATTAAATGGAAGCGTGGTCCACATCCATTAACAC-- 1210  
Db 720 AAC-CGGAGAGATTCGCTATTTAATTAATGGAAGGATGGTCCCATATCCACAACACCT 778  
QY 1211 TTGGGAACCTGAGAAACGC- - - - TGNAGCAACAATAATGTTAAAGGAATGAACAACACTGG 1266  
Db 779 TGGGGACAGAGAAGAACCCCTCAACGACAGCAAAATGTTAAGAGGGAATGAAAAAATGG 838  
QY 1267 ACAACT--ACAAGAAAAGGATCAGGACACAAAGCGTGGCTGCAAAAATGCTTCTCCAGA 1324  
Db 839 ATAAATTATAGAAAAAGAAATCGGNAACCAAGATGGGTGGAAAAAGGCCTCTCCC 898  
QY 1325 AGATGTGG 1332  
Db 899 AAAAGAGG 906

RESULT 6  
AUI17528 899 bp mRNA linear EST 19-OCT-2000  
LOCUS AUI17528 HEMBA1 Homo sapiens cDNA clone HEMBA1001604 5', mRNA  
DEFINITION sequence.  
ACCESSION AUI17528  
VERSION AUI17528.1 GI:10932503  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 899)  
AUTHORS Ota.T., Nishikawa.T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,  
Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and  
Isogai,T.  
TITLE HRI human cDNA project  
JOURNAL Unpublished (2000)  
COMMENT Contact: Takao Isogai  
Genomics Laboratory  
Helix Research Institute  
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan  
Tel: 81-438-52-3951  
Fax: 81-438-52-3952  
Email: genomics@hri.co.jp  
HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix  
Research Institute; cDNA library construction: Department of  
Virology, Institute of Medical Science, University of Tokyo, and  
Helix Research Institute.  
FEATURES  
Source Location/Qualifiers  
1..899  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="HEMBA1001604"  
/clone\_lib="HEMBA1"  
/tissue\_type="whole embryo, mainly head"  
/dev\_stage="embryo, 10 weeks"  
/note="Vector: pME18SFL3"  
BASE COUNT 275 a 182 c 221 g 216 t 5 others  
ORIGIN

Query Match 6.9%; Score 458.4; DB 9; Length 899;  
Best Local Similarity 71.6%; Pred. No. 1.2e-53;  
Matches 639; Conservative 0; Mismatches 250; Indels 3; Gaps 3;  
QY 2276 TCACAAGAGCTGAACCATTTTTTACTAAGAAGATGTAAGAAAGATGTAAGAAAGCTTT 2335  
Db 1 TCATAGGTGCTAGAGCCTTCTCTCCGGAGAGTCAAAAAGATGTGGAGAAATCCCT 60

QY 2336 ACCTGCTAAGGTTGAACAAATTCCTGAGGATGGAAATGAGTGCATTCGCAAGCAATATTA 2395  
Db 61 TCCTGCTAAAAGTGAACAGATTCTCAGGTTGGAGATGTCAGCCCTTCAGAAACAGATTTA 120  
QY 2396 CAAGTGGATTTTAAACAAGGAATTTAAAGCCCTCAGTAAAGGTTCAAAGGTCAGTACCTC 2455  
Db 121 CAAAGTGGATTTCTCAGCAGGAATTTACAAGGCTCTTGCAAGAGCAACAGAGGAGCAGCATC 180  
QY 2456 AGGCTTTCTGAACATTATGATGAAGCTTAAGAAGTCTGTGAACCATTTGCTACCTCATTA 2515  
Db 181 TGGTTTTCTTAATTTGATGAAGACTGAAAAATGTTGCAACCACTGCTATCTGATTA 240  
QY 2516 GCACACAGATGATAATGAATTTCTATATAACAGGAGGCCCTTACAGCATTTGATAGTAG 2575  
Db 241 ACCCCCTGAAGAAATGAAAGGGAATGACAGAGAGATTCCTTCTGCTCCCTCATAGGAG 300  
QY 2576 CAGCGGAAACTAATCCTTCTTGACAAGCTACTGATTTGCTCGGAGAACGTTGGCAACAG 2635  
Db 301 CAGTGGGAAGTTGATTTTATTAGCAAACTGTTGACAAGACTTCGAGAAAGGGGAATCG 360  
QY 2636 AGTCTGATTTCTCTCAGATGGTGAAGTGTGGACATCCTAGCAGACATATCTGAAGTA 2695  
Db 361 AGTCTTATCTCTCAGATGGTGAAGTGTGGATATCCTGGCTGAATACCTACTAT 420  
QY 2696 TCGCCAGTTTCCCTCCAGAGACTTTGATGGATCAATAAAGGGGAATTTGAGGAAGCAAGC 2755  
Db 421 TAAACACTATCCTTTCCAGGCTCTGGATGGTCCATCAAGGGAGAAATCCGAAACAGGC 480  
QY 2756 ACTGGATCATTTCAATGCAAGAGATCAGAGGATTTCTGTTTTTACTGCTACAGAGC 2815  
Db 481 ACTGGACCACCTTCAATGCAGATGGGTCTGAGGACTTCTGTTTCTGCTCGACAAGGGC 540  
QY 2816 TGGAGGATTAAGTATTAACTTGGCATCTGCTGACACTGTAGTATTATTTGATTTCTGACTG 2875  
Db 541 TGTGTCCTGGGAATCAATTTGGCTTCAGCGGACACAGTCTCATCTTTTACTTCTGACTG 600  
QY 2876 GAATCCACAGAAATGATCTGCAGGCACAGCGAGAGTCTATAGAAATTTGGACAGAGAAACA 2935  
Db 601 GAACCCCAAGATGACTTGCAGGCACAAAGCCGCGCATAGAATTTGGTCAGAAGAAGCA 660  
QY 2936 GGTTAATATTTATCGGCTAGTACAAAGAGTACAGTAGAAGAGATATTTCTTGAAGAGC 2995  
Db 661 GGTAAATATTTACCGCTTAGTTACAAAGGGGACTGTGGAGGANGAGATCATAGAAGC-GC 719  
QY 2996 CAAGAAGAAGATGGTCTAGACCATTTAGTAATTCAGAGAAATGGACACGACAGAAAAAC 3055  
Db 720 CAAAAGAAGATGGTATTAGATCATCTGGTGATTCANCGATGGACACCACTGCCGGAC 779  
QY 3056 TGTTCGCATACAGGTTCAACTCCATCCATCAAGCTCTACACCTTTTAAATAAGAGAGTTATC 3115  
Db 780 GATCCTGGAAAACAACACTCAGGAAGTCCAACCTCAAAATCCTTTTAAATAAGAAAAGCT-GC 838  
QY 3116 AGCTATTTTGAAGTTGGTGGCTGAGGAACCTCTTTTAAAGAACCTGCAAGGAGAA 3167  
Db 839 AGCTTTTGAATTTGG-ACCNAGGATCTNTTCAAAGACTTGAAGGGGAGGA 889

RESULT 7  
AUI25712 866 bp mRNA linear EST 23-OCT-2000  
LOCUS AUI25712 NT2RM4 Homo sapiens cDNA clone NT2RM4002061 5', mRNA  
DEFINITION sequence.  
ACCESSION AUI25712  
VERSION AUI25712  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 866)  
AUTHORS Ota.T., Wakamatsu,A., Ozawa,M., Ishii,S., Saito,K., Yamamoto,J.,  
Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki,Y., Sugano,S. and  
Isogai,T.

TITLE HRI human cDNA project (Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S., Saito,K., Yamamoto,J., Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki,Y., Sugano,S., Isogai,T.)  
JOURNAL Unpublished (2000)  
COMMENT Contact: Takao Isogai  
Genomics Laboratory  
Helix Research Institute  
1532-3 Iana, Kisarazu, Chiba 292-0812, Japan  
Tel: 81-438-52-3951  
Fax: 81-438-52-3952  
Email: genomics@hri.co.jp  
HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix Research Institute; cDNA library construction: Department of Virology, Institute of Medical Science, University of Tokyo, and Helix Research Institute.  
FEATURES Location/Qualifiers  
source 1..866  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="NT2RM402061"  
/clone\_lib="NT2RM4"  
/cell\_type="teratocarcinoma"  
/cell\_line="NT2"  
/note="Vector: pME18SFL3; mRNA from uninduced NT2 neuronal precursor cells"  
BASE COUNT 312 a 149 c 196 g 207 t 2 others  
ORIGIN

Query Match 6.9%; Score 456.6; DB 9; Length 866;  
Best Local Similarity 81.3%; Pred. No. 2.1e-53;  
Matches 568; Conservative 0; Mismatches 119; Indels 10; Gaps 3;  
QY 3631 ACAGAAATTTGGTGGCCCTCTGGAAGGTTAGATGCTGTAGCTAGATCTGGAAGTGG 3690  
Db 1 ATAAAGAAATTTGGTGGCCCTCTGGAAGGTTAGATGCTGTAGCTAGATCTGGAAGTGG 60  
QY 3691 TTGATAAATCTGAGACAGACCTTTAGACGCTTTGGGTGAACCTTTACATATGATGTCATTA 3750  
Db 61 TTGATAGTCAGAAACAGACCTTTAGACGCTTTGGGTGAACCTTTACATATGATGTCATTA 120  
QY 3751 AGGCTTTAAAGGACAAATCTATCTGGAAGGTTAGATGCTGTAGCTAGATCTGGAAGTGG 3810  
Db 121 AAGCATTTAAAGGATAGTTCTTTCAGGAACAGAACAGTGTGTAGCTGGAAGTGG 180  
QY 3811 AAGGCCCAACGTTTCGAATCTCAGGAGTGCAGGTGAATGCAGAACTAGTCTCTCAGC 3870  
Db 181 AGGCTCAACATTTCCGAATATCAGGAGTGCAGGTGAATGCAGAACTAGTCTCTCAGC 240  
QY 3871 AAGAAGAGCTGGCCACCACTGCACAAATCCATCTTCAGATCCAGAACAGAAAGAT 3930  
Db 241 AAGAAGANTATACCTTTGCACAAATCCATCTTCAGATCCAGAACAGAAAGAT 300  
QY 3931 ATGTCATCCATGTCACACCAAGCTGCTCCTTCATGATATAGATTTGGGTGAAGAGATG 3990  
Db 301 ATATATPCCATGTCACACCAAGGAGCTGCTCCTTCATGATATAGATTTGGGTGAAGAGATG 360  
QY 3991 ATTCCAATCTGTTAGTGGCATCTATGAATATGCTATGCGAGCTGGGAATGATAAAAA 4050  
Db 361 ATTCCAATTTGTTAATTTGGCATCTATGAATATGCTATGCGAGCTGGGAATGATAAAAA 420  
QY 4051 TGGATCCAGATCTAGCTTTAACACAGAGATTTTACCTGATGATCCAGACAGAAACCC 4110  
Db 421 TGGATCCAGCTACCTTAAACACAGAGATTTTACCTGATGATCCAGACAGAAACCC 480  
QY 4111 AGGCAAGAGCTGACAGACCGTGCAGACTACCTATTAAATTTAGTAAATAGACCTTG 4170  
Db 481 AAGCAAAACAGTTGCAGACCGTGCAGACTACCTATTAAATTTAGTAAATAGACCTTG 540  
QY 4171 CAAGAAGAGACAGCAAGAGCTTGTGCTGTCAGCAATTCACAGAGAGGAGACAGAA 4230  
Db 541 CAAAAAAGAGC-----TCTTCTGGTGGGAGCTTCAAGAGGAGAGAAAGCAAGAG 594

QY 4231 ATAGAAGAATRA---GATGAAGCTTCAAAATTAAGAAGAAATTAAGAAGATGATCTTT 4287  
Db 595 CTAGAAGAAATTAAGCAATGAAGTCTATATAAAGTGAAGAGGAATTAAGAAGATGATCTTT 654  
QY 4288 CACCAACACCTTCAGAAAATCTGTAGAGATGATGA 4324  
Db 655 C-TTCTGCGCTTCAGAGAAGTCTGTAGAGATGATGA 690  
RESULT 8  
AW997058/c  
LOCUS  
DEFINITION QV3-BN0047-150400-152-c03 BN0047 Homo sapiens cDNA, mRNA sequence.  
ACCESSION AW997058  
VERSION AW997058.1 GI:8257292  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 686)  
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., de Oliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,K.R., Reis,D.F., de Souza,S.J. and Simpson,A.J.  
Shotgun sequencing of the human transcriptome with ORF expressed sequence tags  
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
JOURNAL MEDLINE 20202663  
COMMENT Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?cl=et2-QV3-BN0047-150400-152-c03&t3=2000-04-15&t4=1)  
Seq primer: puc 18 forward  
High quality sequence start: 19  
High quality sequence stop: 678.  
FEATURES Location/Qualifiers  
source 1..686  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="BN0047"  
/dev\_stage="Adult"  
/note="Organ: breast\_normal; Vector: puc18; Site\_1: Sma1; Site\_2: Sma1; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 165 a 154 c 126 g 241 t  
ORIGIN  
Query Match 6.9%; Score 452.8; DB 9; Length 686;  
Best Local Similarity 82.6%; Pred. No. 7.8e-53;  
Matches 542; Conservative 0; Mismatches 112; Indels 2; Gaps 2;  
QY 3522 TCCATCACAAAGAAAGAGCGCCAAAAGCGTGGAGACCTGCACCATTCCT-CCAGA 3580  
Db 686 TCCATCTCAAGAGGAAAGCGCCAAAAGAGCGGAGGACCGGATATCTCCCGGA 627  
QY 3581 AAATATAAGGATTTAGTGTATGCAGATACAGGCGGTTTATCAAGAGTTACAGAAAT 3640  
|||||

```
Db 626 GAATATTAAGGATTTAGTGATCGAAGAAATTACCCCTTTTATCAAGAGCTATTAAGAAATT 567
QY 3641 TGTGCGCCCTCTCGAAAGGTTAGATGCTGTAGCTAGAGATGCTGAACCTGTTGATAAATC 3700
Db 566 TGTGTTCTCTCGAAGATTAGATGCAATTTGTCAGATGCTGTAGTTAGTTGATAGTC 507
QY 3701 TGACAGACCTTTAGACGTTTGGGTGAACCTTGACATAATGATGATGATTAAGGCTTTAAA 3760
Db 506 AGAAACAGACCTTTAGATGACTGGGGAATTTGGTACATAATGTTGTCATTAAGACATTA 447
QY 3761 GGCAATTCATCTGGACAAGACAGAGAGGTAGACTTGGGAAAGTTAAAGGCCCAAC 3820
Db 446 GGATAGTTCTTCAGGAACAGCAAGAGGTGTAGACTCGGAAAGAGTGAAGGGTCCAAC 387
QY 3821 GTTTCGAATCTCAGGAGTGCAGGTGAATGCANAACCTAGTATCATCTCTCAGCAAGAAAGAGCT 3880
Db 386 ATTCGGAATATCAGGAGTACAGGTGAATGCCAACTAGTATCATCTCCCATGAAGAAAT 327
QY 3881 GGCACCAGCTGCACAAATCCATCTCTCAGATCCAGAAAGAAAGAAAGATATGTCATCCC 3940
Db 326 AATACCTTTGCACAAATCCATCTCTGATCCAGAAAGAAAGAAAGCAGTATATCTATCCC 267
QY 3941 ATGCCACACCAAGCTGCCTCACTTCGATATAGATTGGGTAAAGAGAGATTCCTCAATCT 4000
Db 266 ATGCCACACCAAGAGCTCATTTTATATAGACTGGGGCAAGAAAGATGATTCCTCAAT 207
QY 4001 GTTAGTAGGCATCTATGAATATGCTATGCTGATGCTGGGAAATGATAAAATGATCCAGA 4060
Db 206 GTTAATGGCATCTATGAATATGATATGAACTGGGAAATGATTAATAATGATCCTGA 147
QY 4061 TCTCAGCTTAACACAGAAATTTTACCTGATGATCCAGACAAAGAAACCCAGCAAGCA 4120
Db 146 CCTCAGCTTAACACAGAAATTTTCCAGATGATCCGATAA-AAACCAACAGCAAAACA 88
QY 4121 GCTACAGACCCGTCAGACTACCTCATTAATTAATCAATTAAGACCTTCGAAAG 4176
Db 87 GTTGACAGCCCGTCAGACTACCTCATCAATTAATTAATTAAGAGATCTTGCAAAA 32

RESULT 9
AW467566/c 609 bp mRNA linear EST 24-FEB-2000
LOCUS he21e07.x1 NCI_CGAP_CML1 Homo sapiens cDNA clone IMAGE:2919684 3'
DEFINITION similar to SW:CHD1_HUMAN O14646 CHROMODOMAIN-HELICASE-DNA-BINDING
PROTEIN 1 ; mRNA sequence.
ACCESSION AW467566
VERSION AW467566.1 GI:7037672
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 609)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapsb@mail.nih.gov
Tissue Procurement: Elisabeth Paietta, Jonathan D. Licht, M.D.,
Michael R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: Life
Technologies, Inc. cDNA Library Arrayed by: Christa Prange, The
I.M.A.G.E. Consortium DNA Sequencing by: Washington University
Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Seq primer: -400p from Glbco
High quality sequence stop: 380.
Location/Qualifiers
1. .609
/organism="Homo sapiens"
/db_xref="taxon:9606"
```

```
/clone="IMAGE:2919684"
/clone_lib="NCI_CGAP_CML1"
/tissue_type="myeloid cells, 18 pooled CML cases, BCR/ABL
rearrangement positive, includes both chronic phase and
myeloid blast crisis"
/lab_host="DH10B"
/Note="Organ: whole blood; Vector: pCMV-SPORT6; Site_1:
Sali; Site_2: NotI; Cloned unidirectionally. Primer:
Oligo dt. Library constructed by Life Technologies."
BASE COUNT 157 a 143 c 98 g 211 t
ORIGIN
```

```
Query Match 6.8%; Score 449.2; DB 9; Length 609;
Best Local Similarity 84.6%; Pred. No. 2.6e-52;
Matches 516; Conservative 0; Mismatches 93; Indels 1; Gaps 1;

QY 2778 GATCAGAGGATTTCTGTTTTTACTGCTACAAAGAGCTGGAGGATAGGTATTAACCTG 2837
Db 609 GGATCAAGGATTTTTG-CTTTGCTGTCCACAAAGAGTGGAGGTCTAGGGATTAATCTA 551
QY 2838 GCATCTGCTGACACTGTAGTTATTTTGTCTGACTGGAATCCACAGAAATGATCTGCAG 2897
Db 550 GCCTTTGCTGACACTGTTGTAATATTGATTCGCGATTGGCATCCACAGAAATGATCTCAG 491
QY 2898 GCACAGCGAGAGCTCATAGAATTGGACAGAGAAACAGGTTAATATTTATCGGCTAGTC 2957
Db 490 GCACAGGCTAGAGCCCATCGAATTTGGCAAAACAAACAGGTGAATATTTATCTCTAGTT 431
QY 2958 ACMAAGGATCAGTAGAAGAATATTTCTTGAAGAGCCCAAGAAAGATGGTGTCTAGAC 3017
Db 430 ACAAGGGATCAGTTTGAAGAAGATATTTCTTGAAGGGCGAAAGAGATGGTTTTAGAT 371
QY 3018 CATTTAGTAATTCAGAGAATGGACACGACAGAGAAACCTGTTCTGCATACAGGTTCAACT 3077
Db 370 CATCTTGTAATTCAAAGAATGGACACACTGGGAAGACAGTACTACATACAGGTTCTGCC 311
QY 3078 CCATCAAGCTCTACACCTTTTATAAAGAGAGTTATCATGCTATTTTGAAGTTTGGTGTCT 3137
Db 310 CCATCAAGTTCTACTCTTTCATAAAGAAGAGTTATCATGCCCCATTTTAAAGTTTGGTGTCT 251
QY 3138 GAGGAACCTTTTAAAGAACTGAGGAGAGAGACAGAGCCCGCCAGGAATGGATATAGAT 3197
Db 250 GAAGAATTTTAAAGAACTGAGGAGAGAGAAACAAAGAGCCCGCCAGGAATGGATATAGAT 191
QY 3198 GAAATCTTGAAGAGAGCTGAAACTCGGGAATAGCCAGGTCCTTACCTGCTAGGGGAT 3257
Db 190 GAAATCTTGAAGAGAGCTGAAACTCATGAAATGAACAGGTCCTTAACTGTAGAGAT 131
QY 3258 GAGTTCCTTTACAGTTCAAGGTGGCGAACTTTTCCAAATATGGATGAAGATGATATTGAG 3317
Db 130 GAATTCCTTTCCAGTTCAAGGTGGCGAACTTCTCAAAATATGGATGAGGATGACATTTAG 71
QY 3318 TTGGACCAAGAAAGAAATTCAGAAATTTGGGAAGAAATCATCCAGAAATCCCAACGAGA 3377
Db 70 TTGGACCTTGAAGAAATTTCAAGAAATTTGGGAGGAAATTTATTTCCAGAAATCAAGAGA 11
QY 3378 AGGATAGAGG 3387
Db 10 CGATTAGAAG 1
```

```
RESULT 10
BG538825
LOCUS BG538825
DEFINITION BG538825
ACCESSION BG538825
VERSION BG538825.1 GI:13531058
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
895 bp mRNA linear EST 03-APR-2001
602568268F1 NIH_MGC_77 Homo sapiens cDNA clone IMAGE:4692504 5',
```







1051	Db	GTTTTACAGAGCAGTGGGAGGGTTACATGTGAGACAACTCGGTGCTCTGGGAGATC	1110
6358	QY	TTACAGTGCATTTGTTTAAAGAAGGAAGTGTCTTATAGGTGAACACTTCAAAACCCAGA	6417
1111	Db	CCCAAGTCAATTTGGGTTCAAAGGCGTGGTAGGTATGAACACTTAAATCCAAA-----	1163
6418	QY	TCAGCCAAAGATTCAATCTGAATCCAAATTTGTTTCCCTCTTTAAACATGGGCAATAATGCA	6477
1164	Db	-CGGCCAACAGTCTCTGTAATGCATGCTTTTCC--CTTTTCATGTGGGCAATATGCA	1220
6478	QY	AATGTGCTATGCAGC---AGTTAATAATTTAGAAGATTTGAATGACCTTATTAACAGAA	6534
1221	Db	AACGTGCTATGCAGCCAGGTTAACATTTTGGAGTAACTTGTATGCTTATTATATAAAC	1280
6535	QY	TGTTACAATGCACACTGATTTACATAGATAAATCTTATCTGACAAAATAAATT	6588
1281	Db	TGTTACAGTACACACTGATTTATATAAAACCTTATATATGACAAAATAAATT	1334

RESULT	12
BE612384	
LOCUS	linear
DEFINITION	EST 20-OCT-2000 60145194F1 NTH_MGC_66 Homo sapiens cDNA clone IMAGE:3855779.5,
ACCESSION	BE612384
VERSION	BE612384.1
KEYWORDS	GI:9893981
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	NIH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> .
TITLE	1 (bases 1 to 1100)
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgapbs.rem@nih.gov Tissue Procurement: DCTD/DTF cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <a href="http://image.llnl.gov">http://image.llnl.gov</a> Plate: LLAM9583 row: g column: 12 High quality sequence stop: 700.

BASE COUNT ORIGIN	Technologies. <sup>a</sup>				Query Match	Best Local Similarity	Score 432; DB 10; Length 1100;	Pred. No. 4, 3e-50;	Matches 639; Conservative 0; Mismatches 170; Indels 35; Gaps 7;
	394 a	233 c	218 g	255 t					
QY 4779	AGCAGCAATGTGTAACACATGTATCAGAAATCCAGATGTGGAAAGACTGAGGAGACT	4838							
Db 18	AACGCACTTGAATCTCAGTGATAGAAATCCAGATGTGGAAAGATTAAAGAGAAAT	77							
QY 4839	ACAAACCAATGATAGTAGTAGCAGGAGCAATTTCTTCGTATAGACATTTATPCAAATAC	4898							

Db	78	ACAATTCAGGATGATGACGAGGACAGTATTCTCTGTAGTACACACTTACTCTAGTAC	137
QY	4899	CATGATCATCACAAAGACAGGACATCAGGAGATGCTTTACAAGAAAAGTACTCCAGGAA	4958
Db	138	CATGATCATCATAAAGACCGACATCAGGGAGATCTTCAAAAAGAAAGTGTATCCAGGAA	197
QY	4959	AGGCCATATTCAGGCTTTCAGTAAATGAGAAAGATCAGAGAGACTGGGATCAGTACAAACAG	5018
Db	198	AGACCCATATTCCTTTTATGTAATGGTAAAGACCATCTGTGATGGGATCAGTACACAGCA	257
QY	5019	GACAGCAGATACTACAGTGATAG--TAAACATAGAAAAGTTAGATGACCACAGGAGCAGA	5075
Db	258	GACAGCAGATATTACAGTCAGACAGAGAGAAACACAGAAAACCTGGATGATCAGAGGAGTGA	317
QY	5076	GACCACAGCTCAAACTTGGAAAGAACTTAAAAGACAGCCGGGTGATTCAGATCACCGC	5135
Db	318	GATCACAGCTCAAAATTTGGAAAGAAAGTTTAAAAGATA--GATCTCATTTCTGATCATCGT	374
QY	5136	TCCCATTTACAGACAGGATACACTCAGATCAACCGTTCACCTTCAGAATACAGCCATCAT	5195
Db	375	TCTCACTCAGATCATCTGGTTTACATTCAGACACCGGTCAAGTCTCGAATATACGCCACT	434
QY	5196	AAATCTTTCCAGAGAGATTATAGATACCACCTCAGACTGGCAATGGACACAGAGCTTCTGGT	5255
Db	435	AAATCTTCCAGGATTTAGGTATCATCTCAGACTGGCAATGGACACAGAGCTTCCAGC	494
QY	5256	AGTGGCCGAGGTCACCACTTAGATCAGAGTCTCTTATGGTTCAGATCTCCCTTAGGA	5315
Db	495	AGTGGCCCTTAGGTCACCTCAGATCAGATCTCTTATGGCTC-----	538
QY	5316	CACAGATCTCCATTTGAACTCATCAGATCAAAAAGTACACCTGAACATACATGAGAT	5375
Db	539	--CAGATCTCCATTTGAACTTCAGTTCGACACAAAGTACACCGGAGCTACTCTGGGT	596
QY	5376	AGCCGGAAGACAT-----AACAAAGACTGACATTTCTCTGGACCTTCTTTTAGCCA--TAT	5429
Db	597	AGTGGGAAACATACAAAACACTGATCTCGTCTTTCTGGACATTTCTTTTAGCCATTAT	656
QY	5430	ACATGAACTAACACAGTAAATTCCTTACATGACTTGAAGAGATGGACTGATATTTCTA	5489
Db	657	ATCATAAACCAACACAGTAATTCGCTTACATGACTTGAAGAATATAAACAGAACTTCTA	716
QY	5490	TCAGTAG-CAGTATGTGTTCTCTTTCC-----AGGATGCAAGGCTTATTATCCCAACAGA	5544
Db	717	CCAGGAGCCAGAATGTCACTCTTTCCACAGGATGACAGAGGCCAATATTATCCCAAGGA	776
QY	5545	AGAAAATATTTTGTATTTAAAGTTTATGCTGCATCTGGCTGCAATATGTTGGGCACCTT	5604
Db	777	GACGAAAAATCTCTATAATTAAGGAGTAAATGGCGCTGTGTTCCACACATGTTCGTAA	836
QY	5605	TTTTT 5608	
Db	837	ATTT 840	
RESULT	13		
LOCUS	AV729238	804 bp	linear
DEFINITION	AV729238	HTC Homo sapiens	cdna clone H7CBBH12 5', mRNA sequence.
ACCESSION	AV729238		
VERSION	AV729238.1	GI:10838659	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Euthera; Primates; Catarrhini; Homiidae; Homo.		
	1 (bases 1 to 804)		
	Gu, Y., Peng, Y., Song, H., Huang, Q., Yang, Y., Gao, G., Xiao, H., Xu, X.,		
	Li, N., Qian, B., Liu, F., Qu, J., Gao, X., Cheng, Z., Xu, Z., Zeng, L., Xu,		
	S., Gu, W., Tu, Y., Jia, J., Fu, G., Ren, S., Zhong, M., Lu, G., Hu, R.,		
	Chen, J., Chen, Z., and Han, Z.		
	Homo sapiens cdna HTC clones		
	Unpublished (2000)		
TITLE			
JOURNAL			





Db 541 AGCATTAACCGTAGTCTTCAGGACAGACGACAGGTGGTAAGACTCGGAA 594

Search completed: August 2, 2002, 16:53:13  
Job time: 12274 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 2, 2002, 20:55:27 ; Search time 12674 Seconds  
(without alignments)  
437.553 Million cell updates/sec

Title: US-08-973-363-12

Perfect score: 265

Sequence: 1 GATGAGATTGTTTCAGTGAA.....AAGAAAGTGAAGGAGAGAG 265

Scoring table:

IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.\*

1: gb.pa.\*

2: gb.hcg.\*

3: gb.in.\*

4: gb.om.\*

5: gb.ov.\*

6: gb.pat.\*

7: gb.ph.\*

8: gb.pl.\*

9: gb.pr.\*

10: gb.ro.\*

11: gb.sts.\*

12: gb.sy.\*

13: gb.un.\*

14: gb.vl.\*

15: em.ba.\*

16: em.fun.\*

17: em.hum.\*

18: em.in.\*

19: em.mu.\*

20: em.om.\*

21: em.or.\*

22: em.ov.\*

23: em.pat.\*

24: em.ph.\*

25: em.pl.\*

26: em.ro.\*

27: em.sts.\*

28: em.un.\*

29: em.vl.\*

30: em.htg\_hum.\*

31: em.htg\_inv.\*

32: em.htg\_other.\*

33: em.htgo\_inv.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
------------	-------	-------------	--------	----	-------------

1	265	100.0	265	6	A58693	A58693 Sequence 12
2	253	95.5	6872	5	AF004397	AF004397 Gallus ga
3	189	71.3	101220	9	AC092372	AC092372 Homo sapi
4	189	71.3	134365	9	AC012624	AC012624 Homo sapi
5	189	71.3	143079	2	AC021449	AC021449 Homo sapi
6	189	71.3	145659	2	AC008531	AC008531 Homo sapi
7	189	71.3	193446	2	AC051946	AC051946 Homo sapi
8	189	71.3	219258	9	AC022121	AC022121 Homo sapi
9	189	71.3	276181	2	AC092382	AC092382 Homo sapi
10	121	45.7	137	6	A58694	A58694 Sequence 13
11	108.8	41.4	7218	6	166494	166494 Sequence 14
12	108.4	40.9	179206	9	CNS01DS6	AL121656 BAC sequ
13	103.2	38.9	72386	2	AC102703	AC102703 Mus muscu
14	102.8	38.8	24259	6	AX251447	AX251447 Sequence
15	102.6	38.7	60565	2	AC023852	AC023852 Homo sapi
16	102	38.5	6668	6	AX346599	AX346599 Sequence
17	102	38.5	61155	2	AC074069	AC074069 Homo sapi
18	102	38.5	113515	6	AX347076	AX347076 Sequence
19	102	38.5	146936	2	AC106157	AC106157 Rattus no
20	102	38.5	173585	2	AC021113	AC021113 Homo sapi
21	102	38.5	196502	2	AL3359957	AL3359957 Sequence
22	101.6	38.3	5586	6	AX348391	AX348391 Sequence
23	101.6	38.3	5890	2	AC110050	AC110050 Homo sapi
24	101.4	38.3	6644	6	E23356	E23356 Virus vecto
25	101.4	38.3	7372	6	E23357	E23357 Virus vecto
26	101.4	38.3	7797	6	E23355	E23355 Virus vecto
27	101.4	38.3	7996	6	E23359	E23359 Virus vecto
28	101.4	38.3	83440	2	AC024285	AC024285 Homo sapi
29	101.4	38.3	256673	2	AC087146	AC087146 Mus muscu
30	101.2	38.2	5387	6	AX344633	AX344633 Sequence
31	101.2	38.2	64767	2	AC102701	AC102701 Mus muscu
32	101.2	38.2	70511	2	AC091083	AC091083 Homo sapi
33	101.2	38.2	231972	2	AC068055	AC068055 Homo sapi
34	101.2	38.2	349980	6	AX344566	AX344566 Sequence
35	100.8	38.0	761	11	CNS06KKG	AL402998 T7 end of
36	100.8	38.0	6171	6	AX345690	AX345690 Sequence
37	100.6	38.0	353	11	CNS06EJ3	AL395173 T7 end of
38	100.6	38.0	587	9	HS337540	AJ337540 Homo sapi
39	100.6	38.0	1926	9	HS802868	AL512733 Homo sapi
40	100.6	38.0	16633	6	AX344576	AX344576 Sequence
41	100.6	38.0	61940	2	AC109312	AC109312 Homo sapi
42	100.6	38.0	159459	2	AC094999	AC094999 Rattus no
43	100.6	38.0	174472	2	AC094034	AC094034 Rattus no
44	100.4	37.9	16033	6	AX346306	AX346306 Sequence
45	100.4	37.9	17934	6	AX346621	AX346621 Sequence

ALIGNMENTS

RESULT	1	A58693	Sequence 12 from Patent WO9639505.	265 bp	DNA	linear	PAT 06-MAR-1998
LOCUS		A58693					
DEFINITION		A58693					
ACCESSION		A58693					
VERSION		A58693.1	GI:3714251				
KEYWORDS		unidentified.					
SOURCE		unclassified.					
ORGANISM		unclassified.					
REFERENCE		1 (bases 1 to 265)					
AUTHORS		Griffiths,R. and Tiwari,B.					
TITLE		AVIAN GHG GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN BIRDS					
JOURNAL		Patent: WO 9639505-A 12 12-DEC-1996;					
COMMENT		ISIS INNOVATION (GB)					
FEATURES		Other Publication AU 5906996 961224.					
source		Location/Qualifiers					
		1..265					
		/organism="unidentified"					
		/db_xref="taxon:32644"					
BASE COUNT		158 a	61 g	30 t			
ORIGIN							

```
Query Match 100.0%; Score 265; DB 6; Length 265;
Best Local Similarity 100.0%; Pred. No. 1.1e-27;
Matches 265; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GATGAGATTGTTTCAGTGAACATCTACATAAAAAATAAAAAACAGAAAAAGAAATGA 60
|||||
Db 1 GATGAGATTGTTTCAGTGAACATCTACATAAAAAATAAAAAACAGAAAAAGAAATGA 60
|||||

Qy 61 AGAAAAAGCCTGAGCAGATATGGTATATAAGAGGAGCTGAAGAAAAAGAGACAAA 120
|||||
Db 61 AGAAAAAGCCTGAGCAGATATGGTATATAAGAGGAGCTGAAGAAAAAGAGACAAA 120
|||||

Qy 121 AGAAGAGAAAAATAAAGGGAATTTGAAGAGGAGAAAAAGAAAAAGAGATAGAAGA 180
|||||
Db 121 AGAAGAGAAAAATAAAGGGAATTTGAAGAGGAGAAAAAGAAAAAGAGATAGAAGA 180
|||||

Qy 181 ATTTAAAGAAAAAGATATATAAGAAAAAGAGAAAAACAAAAGTAAAAAGATCCACACAGAA 240
|||||
Db 181 ATTTAAAGAAAAAGATATATAAGAAAAAGAGAAAAACAAAAGTAAAAAGATCCACACAGAA 240
|||||

Qy 241 AGAAAAAGAACTGAAGAGAGAG 265
|||||
Db 241 AGAAAAAGAACTGAAGAGAGAG 265
|||||

RESULT 2
AF004397 AF004397 6872 bp mRNA linear VRT 08-OCT-1997
LOCUS
DEFINITION Gallus gallus chromo-helicase-DNA-binding on the z chromosome
protein, variant with hydrophilic domain, (CHD-Z) mRNA, complete
cds.
ACCESSION AF004397
VERSION AF004397.1 GI:2501845
KEYWORDS
SOURCE chicken.
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
REFERENCE 1 (bases 1 to 6872)
Griffiths, R. and Korn, R.M.
A CHD1 gene is z chromosome linked in the chicken Gallus domesticus
Gene 197 (1-2), 225-229 (1997)
MEDLINE 97473516
REFERENCE 2 (bases 1 to 6872)
Griffiths, R. and Korn, R.M.
Direct Submission
Submitted (16-MAY-1997) Zoology, Molecular Lab, Glasgow University,
Glasgow G12 8QQ, UK
FEATURES
Location/Qualifiers
source 1..6872
/organism="Gallus gallus"
/db_xref="taxon:9031"
1..6872
gene /gene="CHD-Z"
228..5654
CDS /gene="CHD-Z"
/function="role in chromatin architecture"
/notes="CHD protein with hydrophilic domain"
/codon_start=1
/product="chromo-helicase-DNA-binding on the z chromosome
protein"
/protein_id="AAC60282.1"
/db_xref="GI:2501846"
/translation="MNGHSDSESVRNSGESSKSDSDSGASGSGSGSSGSSDSSG
SQSGSDSGSGSQSGSESSTREKQVQAKPKADGSEFWKSPSILAVQRSAY
LKQOQOQAAASDSDSGSDSDSDSDSKKKKDKEDWMQSGSGSVSGTGS
DSASADGDKSSCEESDSEYEPKNVKSRPKSPRIKPKSGKSTGOKKQLDSEEE
DDDEDYDKGSRQAQVNVYKAEETKTDSDDLVECGEDVPOTEDEFEIEKPMQ
SRIGRKATGASTTIYAVADGPNAGFEKSELGEIQILIKWGMHSHLHWETET
LKQNVKGMKNDLNYKKKQOETRWLKNASPEDVEYINCQQLTDDLHKQYOQIVERII
```

```
AHSNQSAAGYPDYCYCKWQGLPYSECSWEDGALLAKKFOARIDEYFSRNQSKTTPFKD
CKVLKQRPFRVALKQPSYIGGHESLELDYQNLGNLWLAHSCNCSILADEMGLG
KITQISFLNYLFHEHOLYGPFLLRVPLSTLTSWOREIQTWAPOMAVRVYLGDTISRN
RLTHEHMHPOTKRLKFNILLTYETILLKDKSFLGGLNWFICGVDAHRLKNDSDLLY
MRTDEKSHNRLLITGTPLONSLLKELSLHPTMPEKFSWEDFEHEHKGREYGVAS
LHKEDEPFLLRVKVKQVSKLPAKVEQILRMESALQOYIKWILLRNYKALSKSGSG
STSGFLNIMELKCCNCHYLKPPDDNEFYKQEAOLHLIRSSGLILDLKLLIRLR
EGRNVLIFSQVMRLDILAELKYPQFQRDGSIGKELRRQALDHFNAESEDFC
FLLSRAGGLGINLASADVTWIFDSQWPNQDLOAARAHRIQOKKOVNIYRLVTGGS
VEEDILERAQKKWLDLHVIOQMDTGTVLTHTGSTPSSSTPENKLELSAILKFGAEE
LPKEPEGEQEPQEMDIDELIKRAETRENEPGLTVGDELLSQFKVANFSNMDEDDIE
LPEPNSRNWEEIIPESORRIIEERQKELEIYMLPRNRNCAKQISFNSGGRKSR
SRRYSGSDSITERKPKKRGPRPTIPREINIKGFSDAEIRREIKYKFGGLRLD
AVARADLVKSETDLRLGELVHNGCIKALKDNSSQGRAGRLGKVGKPTFRISGV
OVNAKLVI SHEEELAPLHIPSIDPERKRYVTPCHTAAHFDIDWGEDDLSLLVGI
FLYAGGSEWMIKMDPDLSTOKTLPDDPKPOAKOLOTRADYLIKLLKDLARKAQ
RLEGAGNSKRRKTRNKKMKKASKIKEEIKSDSSQPSSEKDEDEEDKNDLIVSVK
HLHKTKTEKNEEPEPDIGKAEERKREKREKREKREKREKREKREKREKREKREK
NKEKREKREKREKREKREKREKREKREKREKREKREKREKREKREKREKREKREK
ESELHQKTSFVCKERMVPKAAKQDRPEKLSREQLHTROCLIIGDHTISEL
KEVTNPEQIKQWRKNLWIFVSKFTFEDARKLHLKLYKHAIKKROESQOHHNDSSNVN
THVIRNPQVRLKETTNDHSDSSRDSYSSDRHLISQYHDDHKRHOQDAYKKSRSKRPY
SAFNSGKHDRMDHYKODSRYSDSKRKLDDHRSRDRSRLNLEGNLKDSEHSHRSH
SDHRIHSDHRSSTSEYSHHRSKSDYRHYSDWQMDHRSASGSPRSLQDQSPYSGRSLP
HRSFPHSHSDHKSTPETHWSSRKT"
misc_feature
4341..4604
/gene="CHD-Z"
/notes="short insert found in longer variant mRNA of CHD-Z"
BASE COUNT 2446 a 1223 c 1520 g 1683 t
ORIGIN
Query Match 95.5%; Score 253; DB 5; Length 6872;
Best Local Similarity 99.6%; Pred. No. 1.9e-26;
Matches 264; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
Qy 1 GATGAGATTGTTTCAGTGAACATCTACATAAAAAATAAAAAACAGAAAAAGAAATGA 60
Db 4341 GATGAGATTGTTTCAGTGAACATCTACATAAAAAATAAAAAACAG-AAAAAGAAATGA 4399
Qy 61 AGAAAAAGCCTGAGCAGATATGGTATATAAGAGGAGCTGAAGAAAAAGAGACAAA 120
Db 4400 AGAAAAAGCCTGAGCAGATATGGTATATAAGAGGAGCTGAAGAAAAAGAGACAAA 4459
Qy 121 AGAAGAGAAAAATAAAGGGAATTTGAAGAGGAGAAAAAGAAAAAGAGATAGAAGA 180
Db 4460 AGAAGAGAAAAATAAAGGGAATTTGAAGAGGAGAAAAAGAAAAAGAGATAGAAGA 4519
Qy 181 ATTTAAAGAAAAAGATATATAAGAAAAAGAGAAAAACAAAAGTAAAAAGATCCACACAGAA 240
Db 4520 ATTTAAAGAAAAAGATATATAAGAAAAAGAGAAAAACAAAAGTAAAAAGATCCACACAGAA 4579
Qy 241 AGAAAAAGAACTGAAGAGAGAG 265
Db 4580 AGAAAAAGAACTGAAGAGAGAG 4604
RESULT 3
AC092372/c AC092372 101220 bp DNA linear PRI 07-DEC-2001
LOCUS Homo sapiens chromosome 5 clone RP11-58M12, complete sequence.
DEFINITION AC092372
ACCESSION AC092372
VERSION AC092372.3 GI:17402768
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 101220)
DOE Joint Genome Institute and Stanford Human Genome Center.
AUTHORS Direct Submission
TITLE Unpublished
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 101220)
```





Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,  
McPheeters,R., Meldrum,J., Meneus,L., Morrow,J., Naylor,J.,  
Norman,C.H., O'Connor,T., O'Donnell,P., Oliver,T.M., Peterson,K.,  
Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,  
Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,  
Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,  
Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,  
Zimmer,A. and Zody,M.

TITLE  
JOURNAL

Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT

On Sep 10, 2000 this sequence version replaced gi:7407963.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence\_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L5154

Center clone name: 58\_M\_12

----- Summary Statistics

Sequencing vector: M13; M7815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 134743 bases at least Q40

Consensus quality: 139227 bases at least Q30

Consensus quality: 140814 bases at least Q20

Insert size: 144000; agarose-fp

Insert size: 142179; sum-of-contigs

Quality coverage: 4.6 in Q20 bases; agarose-fp

Quality coverage: 4.7 in Q20 bases; sum-of-contigs

-----

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 10 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 38820: contig of 38820 bp in length  
\* 38821 38920: gap of 100 bp  
\* 38921 40411: contig of 1491 bp in length  
\* 40412 40511: gap of 100 bp  
\* 40512 43279: contig of 2768 bp in length  
\* 43280 43379: gap of 100 bp  
\* 43380 46905: contig of 3526 bp in length  
\* 46906 47005: gap of 100 bp  
\* 47006 51830: contig of 4825 bp in length  
\* 51831 51930: gap of 100 bp  
\* 51931 62619: contig of 10689 bp in length  
\* 62620 62719: gap of 100 bp  
\* 62720 75408: contig of 12689 bp in length  
\* 75409 75508: gap of 100 bp  
\* 75509 92516: contig of 17008 bp in length  
\* 92517 92616: gap of 100 bp  
\* 92617 106409: contig of 13793 bp in length  
\* 106410 106509: gap of 100 bp  
\* 106510 143079: contig of 36570 bp in length.

FEATURES

source

1. .143079

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_lib="RP11-58M12"

/clone\_lib="RPC1-11 Human Male BAC"

1. .38820

/note="assembly\_fragment"

clone\_end:SP6

vector\_side:left"

38921..40411

misc\_feature

misc\_feature

misc\_feature /note="assembly\_fragment"  
40512..43279  
misc\_feature /note="assembly\_fragment"  
43380..46905  
misc\_feature /note="assembly\_fragment"  
47006..51830  
misc\_feature /note="assembly\_fragment"  
51931..62619  
misc\_feature /note="assembly\_fragment"  
62720..75408  
misc\_feature /note="assembly\_fragment"  
75509..92516  
misc\_feature /note="assembly\_fragment"  
92617..106409  
misc\_feature /note="assembly\_fragment"  
106510..143079  
misc\_feature /note="assembly\_fragment"

clone\_end:T7

vector\_side:right"

BASE COUNT 43971 a 26246 c 26678 g 45278 t 906 others

ORIGIN

Query Match 71.3%; Score 189; DB 2; Length 143079;

Best Local Similarity 84.5%; Pred. No. 4.6e-18;

Matches 224; Conservative 0; Mismatches 40; Indels 1; Gaps 1;

QY 1 GATGAGATTGTTTCAGTGAACATCTACATAAAAAATAAAAAACAGAAAAAGAAATGA 60

Db 117669 GATGAGATCAGTTCTGTGAACATCCAAATAAAAAATTTAAACAGAAAG-AGACAGTGA 117727

QY 61 AGAAAAGCCCTGAGCCAGATATTGTTATAAAGAGGAGCTGAAGAAAAAGAGAGACAAA 120

Db 117728 AGAAAACCTGAGCCAGATGTTTATATAAGAGGACCAAGAAAAAGAGGAGGAGCAAA 117787

QY 121 AGAAGAGGAAATATAAGGGAATTGAAAGGGAGAGAAAAAGAAAAAGAGATTAAGAAGA 180

Db 117788 AGAAAAGGAGATATAAAAAAGAACTTAAAGGGAGATATAAAAAAGAGAGATAAGAAGA 117847

QY 181 ATTAAAGAAAAAGATAATAAGAAAAAGAGAGAAAAACAAGATAAAGAAATCCACACAGAA 240

Db 117848 TATAAGGAAAAAGATTTTAAAGAAAAAGAGAAAAACAAGATAAAGAGCTATACAGAA 117907

QY 241 AGAAAAGAGAGTGAAGGAGAGAGAG 265

Db 117908 AGAAAAGACATAAAGGAGAGAAAG 117932

RESULT 6

AC008531

LOCUS

DEFINITION Homo sapiens chromosome 5 clone CTC-480B11, WORKING DRAFT SEQUENCE,  
7 ordered pieces.

AC008531

AC008531.3 GI:12830078

VERSION HTG; HTGS\_PHASE2; HTGS\_DRAFT.

KEYWORDS

SOURCE human.

ORGANISM

Homo sapiens

Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 145659)

TITLE Sequencing of Human Chromosome 5

JOURNAL Unpublished

2 (bases 1 to 145659)

DOE Joint Genome Institute.

Direct Submission

TITLE

JOURNAL

COMMENT

On Feb 14, 2001 this sequence version replaced gi:7528342.

-----Genome Center

Center: Joint Genome Institute

Center Code: JGI

Web site: http://www.jgi.doe.gov

-----  
Project Information  
Center Project Name: 369535  
Center clone name: CIT-HSPC\_480B11  
-----  
Summary Statistics  
Consensus quality: 139128 bases at least Q40  
Consensus quality: 142556 bases at least Q30  
Consensus quality: 143744 bases at least Q20  
Estimated insert size: 148000; pulse field gel estimation  
Estimated insert size: 145059; sum-of-contigs estimation  
Quality coverage: 6.27 in Q20 bases; pulse field gel estimation  
Quality coverage: 6.4 in Q20 bases; sum-of-contigs estimation.  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 7 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* provided by the submitter.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.  
\* 1 56174: contig of 36174 bp in length  
\* 56175 56274: gap of unknown length  
\* 56275 100874: contig of 44600 bp in length  
\* 100875 100974: gap of unknown length  
\* 100975 113127: contig of 12153 bp in length  
\* 113128 113227: gap of unknown length  
\* 113228 118190: contig of 4963 bp in length  
\* 118191 118290: gap of unknown length  
\* 118291 119694: contig of 1404 bp in length  
\* 119695 119794: gap of unknown length  
\* 119795 123297: contig of 3503 bp in length  
\* 123298 123397: gap of unknown length  
\* 123398 145659: contig of 22262 bp in length.  
FEATURES  
source  
1..145659  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="5"  
/clone="CITC-480B11"  
/clone\_lib="CalTech human BAC library C"  
BASE COUNT 42561 a 26309 c 27580 g 48609 t 500 others  
ORIGIN

Query Match 71.3%; Score 189; DB 2; Length 145659;  
Best Local Similarity 84.5%; Pred. No. 4.6e-18;  
Matches 224; Conservative 0; Mismatches 40; Indels 1; Gaps 1;

QY 1 GATGAGATTGTTTCAGTGAACATCTACATAAAAAATAAAAAACAGAAAAAGAAATGA 60  
||||| ||| ||||||| ||| ||||||| ||||||| ||||||| ||| |||  
Db 48286 GATGAGATCAGTTCGTGAACATCCAAATAAAAAATTAACACAGAAG-AGACAGTGA 48344  
||||| ||| ||||||| ||| ||||||| ||||||| ||||||| ||| |||  
QY 61 AGAAAGCCCTGAGCCAGATATTGTTATAAAGAGGAAGCTGAAGAAAAAGAGAGACAAA 120  
||||| ||| ||||||| ||| ||||||| ||||||| ||||||| ||| |||  
Db 48345 AGAAAGCCCTGAGCCAGATATTGTTATAAAGAGGAAGCAACAGAGAGGGAAGCAAA 48404  
||||| ||| ||||||| ||| ||||||| ||||||| ||||||| ||| |||  
QY 121 AGAGAGGAAAAATAAAGGGAATTGAAAAGGAGAAAAAGAGAGGATAAGAAAGA 180  
||||| ||| ||||||| ||| ||||||| ||||||| ||||||| ||| |||  
Db 48405 AGAAAGGAGATAAAGAGAACTTAAAGGGAGATAAAGAGAGAGGATAAGAAAGA 48464  
||||| ||| ||||||| ||| ||||||| ||||||| ||||||| ||| |||  
QY 181 ATTAAGAAAAAGATAATAAAGAGAGAGAAAAACAAAGTAAAGAACTCCACACAGAA 240  
||||| ||| ||||||| ||| ||||||| ||||||| ||||||| ||| |||  
Db 48465 TATAAGAAAAAGATTTTAAAGAAAAAGAGACAAACAAAGTAAAGAACTATACAGAA 48524  
||||| ||| ||||||| ||| ||||||| ||||||| ||||||| ||| |||  
QY 241 AGAAAGAGAGTCAAGGAGAGAG 265  
||||| ||| ||||||| ||| ||||||| ||||||| ||||||| ||| |||  
Db 48525 AGAAAGACATAAAGGAGAGAG 48549  
||||| ||| ||||||| ||| ||||||| ||||||| ||||||| ||| |||

RESULT 7

AC091946 193446 bp DNA linear HTG 09-JUN-2001  
Homo sapiens chromosome 5 clone RP11-360I2, \*\*\* SEQUENCING IN  
PROGRESS \*\*\*, 33 unordered pieces.  
AC091946  
AC091946.1 GI:14333882  
HTG: HTGS\_PHASE1.  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 193446)  
DOE Joint Genome Institute.  
Sequencing of Human Chromosome 5  
Unpublished  
2 (bases 1 to 193446)  
DOE Joint Genome Institute.  
Direct Submission  
Submitted (09-JUN-2001) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
Center: Joint Genome Institute  
Center Code: JGI  
Web site: http://www.jgi.doe.gov  
-----  
Project Information  
Center Project Name: 544799  
Center clone name: RPCI-11\_360I2  
-----  
Summary Statistics  
Consensus quality: 157767 bases at least Q40  
Consensus quality: 180259 bases at least Q30  
Consensus quality: 184175 bases at least Q20  
Estimated insert size: 204590; agarose-fp estimation  
Estimated insert size: 190246; sum-of-contigs estimation  
Quality coverage: 5.38 in Q20 bases; agarose-fp estimation  
Quality coverage: 5.79 in Q20 bases; sum-of-contigs estimation.  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 33 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1358: contig of 1358 bp in length  
\* 1359 1458: gap of unknown length  
\* 1459 2867: contig of 1209 bp in length  
\* 2868 2767: gap of unknown length  
\* 2768 3823: contig of 1056 bp in length  
\* 3824 3923: gap of unknown length  
\* 3924 5920: contig of 1997 bp in length  
\* 5921 6020: gap of unknown length  
\* 6021 7494: contig of 1474 bp in length  
\* 7495 7594: gap of unknown length  
\* 7595 9096: contig of 1502 bp in length  
\* 9097 9196: gap of unknown length  
\* 9197 10260: contig of 1064 bp in length  
\* 10261 10360: gap of unknown length  
\* 10361 12460: contig of 2100 bp in length  
\* 12461 12560: gap of unknown length  
\* 12561 14611: contig of 2051 bp in length  
\* 14612 14711: gap of unknown length  
\* 14712 16381: contig of 1670 bp in length  
\* 16382 16481: gap of unknown length  
\* 16482 17968: contig of 1487 bp in length  
\* 17969 20434: gap of unknown length  
\* 20435 20534: contig of 2366 bp in length  
\* 20535 23516: gap of unknown length  
\* 23517 23615: contig of 2961 bp in length  
\* 23616 27563: gap of unknown length  
\* 27564 27663: gap of unknown length

AC091946  
LOCUS  
DEFINITION  
AC091946  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT





RESULT	11
LOCUS	I66494/c
DEFINITION	Sequence 14 from patent US 5670367.
ACCESSION	I66494
VERSION	I66494.1 GI:2724471
KEYWORDS	.
SOURCE	Unknown.
ORGANISM	Unclassified.
REFERENCE	1 (bases 1 to 7218)
AUTHORS	Dorner,F., Scheiflinger,F. and Falkner,F.Gunter.
TITLE	Recombinant fowlpox virus
JOURNAL	Patent: US 5670367-A 14 23-SEP-1997;
FEATURES	Location/Qualifiers
source	1..7218
BASE COUNT	1944 a 1491 c 1486 g 1929 t 368 others
ORIGIN	/organism="unknown"
Query Match	41.4%; Score 109.8; DB 6; Length 7218;
Best Local Similarity	3.4%; Pred. No. 6.3e-07;
Matches	9; Conservative 212; Mismatches 44; Indels 0; Gaps 0;
QY	1 GATCAGATTGTTTCAGTGAACATCTACATAAAAAAATAAAACAGAAAAAGAAAATGA 60             :
Db	1448 GAAGAATTGTGTAACRTRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1389 :
QY	61 AGAAAAGCTTGAGCCAGCATATTGGTATAAAGAGGAGAGCTGAAGAAAAAGAGAGACAAA~120 :
Db	1388 RRR 1329 :
QY	121 AGAAGAGAAAATAAAAGGGAATTGAAAGGGAGAAAAAGAAAAAGAGGATAAGAAAGA 180 :
Db	1328 RRR 1269 :
QY	181 ATTTAAAGAAAAGACAATAAAGAAAAGACAGAAAAAGCAAGTAAAGAAATCCACAGAA 240 :
Db	1268 RRR 1209 :
QY	241 AGAAAAGAGAGTGAAGCAAGAGAG 265 :
Db	1208 RRR 1184 :
RESULT	12
LOCUS	CNS01DS6/c
DEFINITION	BAC sequence from the SP4 candidate region at 2p21-2p22 BAC 367K01 of ClTB_978_SKB library from chromosome 2 of Homo sapiens (Human), complete sequence.
ACCESSION	CNS01DS6 179206 bp DNA linear PRI 02-MAR-2000
VERSION	AL121656.2 GI:7159617
KEYWORDS	HTG; HTGS_DRAFT; SP4 genomic DNA interval.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	1 (bases 1 to 179206) Hazan,J., Fonknechten,N., Mavel,D., Paternotte,C., Samson,D., Artiguenave,F., Davoine,C.S., Cruaud,C., Durr,A., Wincker,P., Brottier,P., Cattolico,L., Barbe,V., Burgunder,J.M., Prud'Homme,J.F., Brice,A., Fontaine,B., Heilig,R. and Weissenbach,J.
TITLE	Spastin, a novel AAA protein, is altered in the most frequent form of autosomal dominant spastic paraplegia
JOURNAL	Nat. Genet. (1999) In press
REFERENCE	2 (bases 1 to 179206)
AUTHORS	Genoscope.
TITLE	Direct Submission











Mon Aug 5 11:51:41 2002

us-08-973-363-12.rge

Page 13



\_\_\_\_\_

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 2, 2002, 20:52:39 ; Search time 1205.34 seconds  
(without alignments)  
377.472 Million cell updates/sec

Title: US-08-973-363-12  
Perfect score: 265  
Sequence: 1 GATGAGATGTTTCAGTGAA.....AAGAGTGAGGAGAGAGAG 265

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_032802.\*  
1: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	265	100.0	265	18	AAT42752
2	122.6	46.3	137	18	Chicken CHD-1A ins
3	102.8	38.8	24259	22	AAS46691
c 4	102	38.5	6668	24	AABL33697
c 5	102	38.5	113515	24	ABL34174
6	101.4	38.3	6644	20	AAAX33181
7	101.4	38.3	7372	20	AAAX33182
8	101.4	38.3	7797	20	AAAX33180
9	101.4	38.3	7996	20	AAAX33184

c 10	100.8	38.0	6171	24	ABL32788	Human immune syste
c 11	100.4	37.9	16033	24	ABL33404	Human immune syste
c 12	100.4	37.9	17934	24	ABL33719	Human immune syste
c 13	99.8	37.7	14006	24	ABL33958	Human immune syste
c 14	99.6	37.6	329	22	AAI82679	Human polynucleoti
c 15	99.2	37.4	3586	22	AA563367	Chemically pretrea
c 16	98.8	37.3	7657	22	AA545477	Chemically pretrea
c 17	98.8	37.3	7657	24	ABL34022	Human immune syste
c 18	98	37.0	6485	22	AA546559	Tumour suppressor
c 19	98	37.0	6485	22	ABL33808	Human immune syste
c 20	97.8	36.9	396	22	AAI94862	Human ovarian can
c 21	97.4	36.8	12007	24	ABL32717	Human immune syste
c 22	97.2	36.7	4167	22	ABA08224	Human ovarian and
c 23	97.2	36.7	4167	22	AAS29236	Genomic sequence #
c 24	97.2	36.7	4167	22	AAS30018	Human lung antigen
c 25	97.2	36.7	4167	22	AAI05022	Human reproductive
c 26	97.2	36.7	4167	22	AAI07543	Human reproductive
c 27	97.2	36.7	4167	22	AAS28711	Genomic sequence #
c 28	97.2	36.7	4167	22	AAK68040	Human immune/hacma
c 29	97.2	36.7	4167	22	AAK89382	Human digestive sy
c 30	96	36.2	6338	24	ABL33126	Human immune syste
c 31	95.6	36.1	9539	22	AAS45347	Chemically pretrea
c 32	95.6	36.1	113515	24	ABL34175	Human immune syste
c 33	95.2	35.9	807	22	AAI21312	Human breast cance
c 34	95.2	35.9	5750	22	AAS46708	Tumour suppressor
c 35	95.2	35.9	5750	24	ABL34008	Human immune syste
c 36	94.4	35.6	6183	24	ABL33197	Human immune syste
c 37	94.2	35.5	398	22	AAI88388	Human polynucleoti
c 38	94.2	35.5	420	22	AAI85952	Human polynucleoti
c 39	94	35.5	8342	24	ABL32501	Human immune syste
c 40	93.6	35.3	5845	24	ABL33662	Human immune syste
c 41	93	35.1	16167	24	ABL33083	Human immune syste
c 42	93	35.1	16167	24	ABL34529	Human metastasis a
c 43	93	35.1	56153	22	AAS46794	Tumour suppressor
c 44	92.8	35.0	6161	24	ABL32411	Human immune syste
c 45	91.4	34.5	6668	24	ABL33696	Human immune syste

ALIGNMENTS

RESULT 1  
AAT42752  
ID AAT42752 standard; cdna; 265 BP.  
XX  
AC AAT42752;  
XX  
XX 12-MAR-1997 (first entry)  
DT Chicken CHD-1A insert motif.  
DE  
DE  
XX Bird; sex determination; chromodomain-Helicase-DNA binding 1 Avian;  
KW CHD-W; chromodomain-Helicase-DNA binding on the W chromosome; ss.  
XX  
OS Gallus sp.  
XX  
XX Key Location/Qualifiers  
FH misc\_difference 52  
FT /\*tag= a  
FT /note= "Base 52 disrupts the reading frame for  
FT the translated amino acid sequence given  
FT in Fig 7"  
XX  
XX W09639505-A1.  
XX  
XX 12-DEC-1996.  
XX  
XX 05-JUN-1996; 96WO-GB01341.  
XX  
XX 06-JUN-1995; 95GB-0011439.  
XX  
XX (ISIS-) ISIS INNOVATION LTD.  
XX

```

PI Griffiths R, Tiwari B;
XX WPI; 1997-043127/04.
XX P-PSDB; AAW08144.
XX Avian chromodomain-helicase-DNA binding genes determine sex in
XX birds - used for sex determ. and to control sex of progeny
XX Disclosure; Fig 7; 76pp; English.
XX
XX A composite nucleotide sequence (AAT42752) and putative translation
XX (AAW08144) sequence are provided of a motif that is found spliced to
XX a proportion of chicken CHD-1A clones. The motif is inserted
XX between bases 4327 and 4328 of the CHD-1A composite sequence
XX (AAT42751). None of the 7 CHD-1 clones examined contained the
XX complete motif. There are no splice donor or acceptor sites within
XX the motif suggesting it is a final rather than an intermediary
XX product of splicing. The motif is also found at the 3' end of the
XX CHD-W clone CC14 (see also AAT42753).
XX
XX Sequence 265 BP; 158 A; 16 C; 61 G; 30 T; 0 other;
SQ
Query Match 100.0%; Score 265; DB 18; Length 265;
Best Local Similarity 100.0%; Pred. No. 1.4e-31;
Matches 265; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GATGAGATTGTTTCAGTGAACATCTACATAAAAAATAAAAAACAGAAAAAGAAATGA 60
DB 1 gatgagattgttcagtgaacatctacataaaaaataaaaaacagaaaaaagaatga 60
QY 61 AGAAAGCCGTGAGCCAGATATTGGTATAAGAGAGGAGCTGAAGAAAAAGAGAGACAAA 120
DB 61 agaaaagcctgagccagatattggtataaagaaggaagctgaagaaaaagagacaaa 120
QY 121 AGCAAGGAAAAATAAAGGGAATGTAAAGCGGAGAAAAAAGAAAAAGAGATGAAGA 180
DB 121 agagaaggaaaaataaagggaatgtgaaaggagaaaaaagaaagagataagaaga 180
QY 181 ATTAAGAGAAAAAGATAAATAAGAAAAAGAGAAAAACAAAAATCCACACAGAA 240
DB 181 attaaagaaaaagataataaagaaaaagagaaaaaagaagtaaaagatccacacagaa 240
QY 241 AGAAAGAGAGTGAAGGAAGAGAG 265
DB 241 agaaaagaagtgaaggagaagaag 265
RESULT 2
AAT42753
ID AAT42753 standard; cDNA; 137 BP.
XX
XX AAT42753;
XX
XX 12-MAR-1997 (first entry)
XX
XX Chicken CHD-W clone CC14 3' motif.
XX
XX Bird; sex determination; chromodomain-Helicase-DNA binding 1 Avian;
XX CHD-W; chromodomain-Helicase-DNA binding on the W chromosome; ss.
XX
XX Gallus sp.
XX
XX Key Location/Qualifiers
XX misc_difference 52
XX /*tag= a
XX /*note= base 52 disrupts the reading frame for
XX the translated amino acid sequence given
XX in Fig 7"
XX
XX WO9639505-A1.
XX
XX 12-DEC-1996.
XX
XX Griffiths R, Tiwari B;
XX WPI; 1997-043127/04.
XX P-PSDB; AAW08144.
XX (ISIS-) ISIS INNOVATION LTD.
XX Griffiths R, Tiwari B;
XX WPI; 1997-043127/04.
XX P-PSDB; AAW08145.
XX Avian chromodomain-helicase-DNA binding genes determine sex in
XX birds - used for sex determ. and to control sex of progeny
XX Disclosure; Fig 7; 76pp; English.
XX
XX A composite (incomplete) nucleotide sequence (AAT42753) and putative
XX translation (AAW08145) sequence are provided of a motif that is found
XX spliced to the 3' end of chicken CHD-W clone CC14 (see also AAT42754).
XX There are no splice donor or acceptor sites within the motif
XX suggesting it is a final rather than an intermediary product of
XX splicing. The motif is also found as an insert in some CHD-1A clones
XX (see also AAT42752).
XX
XX Sequence 137 BP; 75 A; 10 C; 34 G; 18 T; 0 other;
SQ
Query Match 46.3%; Score 122.6; DB 18; Length 137;
Best Local Similarity 93.4%; Pred. No. 2.2e-10;
Matches 128; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
QY 1 GATGAGATTGTTTCAGTGAACATCTACATAAAAAATAAAAAACAGAAAAAGAAATGA 60
DB 1 gatgagattgttcagtgaacatccacataaaaaataaaagcagaaaaaagaatga 60
QY 61 AGAAAGCCGTGAGCCAGATATTGGTATAAGAGAGGAGCTGAAGAAAAAGAGAGACAAA 120
DB 61 agaaaagatgagccagagattggtataaagaaggaagctggagaaaaaagagagacaaa 120
QY 121 AGAAGAGGAAAAATAAAA 137
DB 121 agagaaggaaaaataaga 137
RESULT 3
AAS46691/C
ID AAS46691 standard; DNA; 24259 BP.
XX
XX AAS46691;
XX
XX 18-DEC-2001 (first entry)
XX
XX Tumour suppressor gene derived chemically modified sequence #414.
XX
XX Human; tumour suppressor gene; oncogene; antitumour; cytostatic;
XX cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
XX cytosine methylation; ds.
XX
XX Homo sapiens.
XX
XX WO200168912-A2.
XX
XX 20-SEP-2001.
XX
XX 15-MAR-2001; 2001WO-EP02955.
XX
XX 15-MAR-2000; 2000DE-1013847.
XX
XX 06-APR-2000; 2000DE-1019058.
XX
XX 07-APR-2000; 2000DE-1019173.
XX
XX 30-JUN-2000; 2000DE-1032529.
XX
XX 01-SEP-2000; 2000DE-1043826.
XX

```



```
RESULT 5
ABL34174/c
ID ABL34174 standard; DNA; 113515 BP.
XX
XX ABL34174;
AC
XX
XX
DT 26-MAR-2002 (first entry)
XX
XX Human immune system associated gene SEQ ID NO: 2147.
DE
XX
XX Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antiandemic; cytosatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmologic;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW gene; ds.
XX
XX Homo sapiens.
OS
XX WO200200928-A2.
PN
XX
XX 03-JAN-2002.
PD
XX
XX 02-JUL-2001; 2001WO-EP07537.
PF
XX
XX 30-JUN-2000; 2000DE-1032529.
PR
XX 01-SEP-2000; 2000DE-1043826.
PR
XX
XX (EPIG-) EPIGENOMICS AG.
PA
XX
XX Olek A, Piepenbrock C, Berlin K;
PI
XX WPI; 2002-130909/17.
DR
XX
XX Nucleic acid comprising fragment of chemically modified gene, useful
PT for diagnosis and treatment of diseases associated with abnormal
PT cytosine methylation -
PT
XX
XX Claim 1; SEQ ID NO 2147; 32pp + Sequence Listing; German.
PS
XX
XX The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention.
XX
XX Sequence 113515 BP; 31803 A; 1174 C; 24020 G; 56518 T; 0 other;
SQ

Query Match 38.5%; Score 102; DB 24; Length 113515;
Best Local Similarity 63.4%; Pred. No. 1.1e-07;
Matches 156; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

QY 19 AACATCTACATAAAATAAAACAGAAAAAGAAATGAGAAAGCCTGAGCCAGA 78
||||| ||| ||||||| ||||| ||||||| ||||| ||||| ||||| ||||| |||||
DB 51723 AACAAATTAACAAAAAATAAAATAAAATAAAATAAAATAAAATAAAATAAA 51664
||||| ||| ||||||| ||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 79 TATTGGTATTAAGAGGAGCTGAAGAAAGAGAGACAGACAAAGAGAGAAATATAAG 138
||||| ||| ||||||| ||||| ||||||| ||||||| ||||||| ||||||| |||||||
DB 51663 AAAAAAATAAAATAAAATAAAATAAAATAAAATAAAATAAAATAAAATAAA 51604
||||| ||| ||||||| ||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 139 GGAATTGAAGGAGAAAAAGAAAAAGAGGATATAAGAAAGATTAAGAAAAAGATAA 198
||||| ||| ||||||| ||||| ||||||| ||||||| ||||||| ||||||| |||||||
DB 51603 AAAAAAATAAAATAAAATAAAATAAAATAAAATAAAATAAAATAAAATAAA 51544
||||| ||| ||||||| ||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 199 TAAAGAAAGAGAGAAACAACTAAAGATCCACAGAAAGAAAGAAAGTGAAGCA 258
||||| ||| ||||||| ||||| ||||||| ||||||| ||||||| ||||||| |||||||
```

```
Db 51543 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 51484
QY 259 AGAGAA 264
| | | |
Db 51483 AAAAAA 51478

RESULT 6
AAX33181
ID AAX33181 standard; DNA; 6644 BP.
XX
XX AAX33181;
AC
XX
XX 25-JUN-1999 (first entry)
DT
XX
XX Base sequence of the plasmid prx-ires-bsr.
DE
XX
XX Cowpox virus; bsr; viral vector; expression; apoptosis; resistance;
KW crmA; bcl-2; bcl-x1; FLIP; survivin; IAP; adenovirus; cancer;
KW autoimmune disease; graft rejection reaction; inflammation;
KW inflammatory disease; ss.
XX
XX Synthetic.
OS
XX Cowpox virus.
PN
XX WO9913073-A2.
XX
XX 18-MAR-1999.
PD
XX
XX 07-SEP-1998; 98WO-JP04010.
PF
XX
XX 08-SEP-1997; 97JP-0259235.
PR
XX
XX (RPRG-) RPR GENCELL ASIA PACIFIC INC.
PA
XX
XX Hamada H;
PI
XX
XX WPI; 1999-243728/20.
DR
XX
XX New apoptosis-resistant virus-sensitive cell
PT
XX
XX Example 1; Page 38-41; 5lpp; English.
PS
XX
XX The present invention describes an apoptosis-resistant virus-sensitive
CC cell line into which an apoptosis resistance gene has been introduced.
CC The recombinant viruses generated are capable of expressing apoptosis-
CC associated genes. These can then be used in a variety of diseases for
CC which the induction of apoptosis by gene transfer, or where the
CC inhibition of harmful apoptosis, is therapeutic. The recombinant viruses
CC are useful as vectors for gene therapy which can be applied to cancer
CC therapy for destroying cancer cells selectively, the treatment of
CC autoimmune diseases and graft rejection reaction, and apoptosis induction
CC therapy for inflammatory cells in inflammatory diseases. Prior arts have
CC encountered the problem where if an adenovirus vector capable of
CC expressing an apoptosis-associated gene is introduced into animal cells,
CC the cells producing the virus will be destroyed because the period of
CC time required to induce cell death by apoptosis is shorter than that
CC required to replicate and produce the virus, resulting in failure to
CC obtain a recombinant virus having the integrated apoptosis-associated
CC gene. In this invention an apoptosis-resistant 293 cell line (having an
CC apoptosis resistant gene introduced) is established and overcomes the
CC problem. The present sequence represents the base sequence of the
CC plasmid prx-ires-bsr, which contains the cowpox virus bsr gene, and
CC is used in an example from the present invention.
XX
XX Sequence 6644 BP; 2166 A; 1573 C; 1424 G; 1481 T; 0 other;
SQ

Query Match 38.3%; Score 101.4; DB 20; Length 6644;
Best Local Similarity 62.4%; Pred. No. 1.9e-07;
Matches 159; Conservative 0; Mismatches 96; Indels 0; Gaps 0;

QY 10 GTTTCAGTGAACATCTACATAAAAAAAAAATAAAACAGAAAAAGAAATGAAGAAAGCC 69
```





CC inhibition of harmful apoptosis, is therapeutic. The recombinant viruses  
CC are useful as vectors for gene therapy which can be applied to cancer  
CC therapy for destroying cancer cells selectively, the treatment of  
CC autoimmune diseases and graft rejection reaction, and apoptosis induction  
CC therapy for inflammatory cells in inflammatory diseases. Prior arts have  
CC encountered the problem where if an adenovirus vector capable of  
CC expressing an apoptosis-associated gene is introduced into animal cells,  
CC the cells producing the virus will be destroyed because the period of  
CC time required to induce cell death by apoptosis is shorter than that  
CC required to replicate and produce the virus, resulting in failure to  
CC obtain a recombinant virus having the integrated apoptosis-associated  
CC gene. In this invention an apoptosis-resistant 293 cell line (having an  
CC apoptosis resistant gene introduced) is established and overcomes the  
CC problem. The present sequence represents the cowpox virus bsr gene which  
CC is used in an example from the present invention.  
XX  
SQ Sequence 7797 BP; 2542 A; 1760 C; 1656 G; 1839 T; 0 other;

Query Match 38.3%; Score 101.4; DB 20; Length 7797;  
Best Local Similarity 62.4%; Pred. No. 1.9e-07;  
Matches 159; Conservative 0; Mismatches 96; Indels 0; Gaps 0;  
QY 10 GTTTCAGTGAACATCTACATAAAATAAATAACACAGAAAAGAAATCAAGAAAGCC 69  
Db 4887 gtgcacaaataaaataaaataaaataaaataaaataaaataaaataaaataaa 4946  
QY 70 TGAGCCAGATATTGGTATAAAGGAAGCTGAAGAAAAGAGAGACAAAGAGAGGA 129  
Db 4947 aaaaaataaaataaaataaaataaaataaaataaaataaaataaaataaaata 5006  
QY 130 AATAAAGGGAATGTAAAGGGAGAGAAAAGAAAAGAGGATAGAAAGAAATTAAGA 189  
Db 5007 aaaaaataaaataaaataaaataaaataaaataaaataaaataaaataaaata 5066  
QY 190 AAAAGTAATAAGAAAGAGAGAAAACAAAGTAAAGATCCACACAGAAAGAAAGA 249  
Db 5067 aaaaaataaaataaaataaaataaaataaaataaaataaaataaaataaaata 5126  
QY 250 AGTGAAGGAAGAGAA 264  
Db 5127 aaaaaataaaataaa 5141

RESULT 9  
AAX33184  
ID AAX33184 standard; DNA; 7996 BP.  
XX  
AC AAX33184;  
XX  
XX 25-JUN-1999 (first entry)  
DT  
XX Base sequence of the plasmid prx-Bcl 2-i-hcd 25.  
DE  
XX Cowpox virus; bsr; viral vector; expression; apoptosis; resistance;  
KW crmA; bcl-2; bcl-x1; FLIP; survivin; IAP; ILP; adenovirus; cancer;  
KW autoimmune disease; graft rejection reaction; inflammation;  
KW inflammatory disease; ss.  
XX  
OS Synthetic.  
OS Homo sapiens.  
XX  
PN WO9913073-A2.  
XX  
XX 18-MAR-1999.  
PD  
XX 07-SEP-1998; 98WO-JP04010.  
PF  
XX 08-SEP-1997; 97JP-0259235.  
PR  
XX (RPRG-) RPR GENCELL ASIA PACIFIC INC.  
PA  
XX Hamada H;  
PI

XX WPI; 1999-243728/20.  
DR New apoptosis-resistant virus-sensitive cell  
XX  
PT Example 3; Page 46-49; Slpp; English.  
XX  
CC The present invention describes an apoptosis-resistant virus-sensitive  
CC cell line into which an apoptosis resistance gene has been introduced.  
CC The recombinant viruses generated are capable of expressing apoptosis-  
CC associated genes. These can then be used in a variety of diseases for  
CC which the induction of apoptosis by gene transfer, or where the  
CC inhibition of harmful apoptosis, is therapeutic. The recombinant viruses  
CC are useful as vectors for gene therapy which can be applied to cancer  
CC therapy for destroying cancer cells selectively, the treatment of  
CC autoimmune diseases and graft rejection reaction, and apoptosis induction  
CC therapy for inflammatory cells in inflammatory diseases. Prior arts have  
CC encountered the problem where if an adenovirus vector capable of  
CC expressing an apoptosis-associated gene is introduced into animal cells,  
CC the cells producing the virus will be destroyed because the period of  
CC time required to induce cell death by apoptosis is shorter than that  
CC required to replicate and produce the virus, resulting in failure to  
CC obtain a recombinant virus having the integrated apoptosis-associated  
CC gene. In this invention an apoptosis-resistant 293 cell line (having an  
CC apoptosis resistant gene introduced) is established and overcomes the  
CC problem. The present sequence represents the base sequence of the  
CC plasmid prx-Bcl 2-i-hcd 25, which contains the human Bcl-2 gene, and  
CC is used in an example from the present invention.  
XX  
SQ Sequence 7996 BP; 2463 A; 2015 C; 1829 G; 1689 T; 0 other;

Query Match 38.3%; Score 101.4; DB 20; Length 7996;  
Best Local Similarity 62.4%; Pred. No. 1.9e-07;  
Matches 159; Conservative 0; Mismatches 96; Indels 0; Gaps 0;  
QY 10 GTTTCAGTGAACATCTACATAAAATAAATAACACAGAAAAGAAATCAAGAAAGCC 69  
Db 5086 gtgcacaaataaaataaaataaaataaaataaaataaaataaaataaaataaa 5145  
QY 70 TGAGCCAGATATTGGTATAAAGGAAGCTGAAGAAAAGAGAGACAAAGAGAGGA 129  
Db 5146 aaaaaataaaataaaataaaataaaataaaataaaataaaataaaataaaata 5205  
QY 130 AATAAAGGGAATGTAAAGGGAGAGAAAAGAAAAGAGGATAGAAAGAAATTAAGA 189  
Db 5206 aaaaaataaaataaaataaaataaaataaaataaaataaaataaaataaaata 5265  
QY 190 AAAAGTAATAAGAAAGAGAGAAAACAAAGTAAAGATCCACACAGAAAGAAAGA 249  
Db 5266 aaaaaataaaataaaataaaataaaataaaataaaataaaataaaataaaata 5325  
QY 250 AGTGAAGGAAGAGAA 264  
Db 5326 aaaaaataaaataaa 5340

RESULT 10  
ABL32788/c  
ID ABL32788 standard; DNA; 6171 BP.  
XX  
AC ABL32788;  
XX  
XX 26-MAR-2002 (first entry)  
DT  
XX Human immune system associated gene SEQ ID NO: 761.  
DE  
XX Human; immune system disease; cytosine methylation; antiasthmatic;  
KW antiarteriosclerotic; anti-naemic; cytosolic; nontropic;  
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;  
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;  
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;  
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;

KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;  
KW gene; ds.  
OS Homo sapiens.  
XX WO200200928-A2.  
XX 03-JAN-2002.  
XX 02-JUL-2001; 2001WO-EP07537.  
XX 30-JUN-2000; 2000DE-1032529.  
XX 01-SEP-2000; 2000DE-1043826.  
XX (EPIG-) EPIGENOMICS AG.  
XX Olek A, Piepenbrock C, Berlin K;  
XX WPI; 2002-130909/17.  
XX Nucleic acid comprising fragment of chemically modified gene, useful  
PT for diagnosis and treatment of diseases associated with abnormal  
PT cytosine methylation  
XX Claim 1; SEQ ID NO 761; 32pp + Sequence Listing; German.  
XX The present invention provides a number of human immune system associated  
CC genes which are modified by the methylation of cytosines. The sequences  
CC can be used in the diagnosis and treatment of immune system disorders,  
CC including eye diseases such as retinopathy, neovascular glaucoma and  
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
CC diseases. The present sequence is a gene of the invention.  
XX Sequence 6171 BP; 1614 A; 118 C; 1339 G; 3100 T; 0 other;  
Query Match 38.0%; Score 100.8; DB 24; Length 6171;  
Best Local Similarity 62.1%; Pred. No. 2.4e-07;  
Matches 159; Conservative 0; Mismatches 97; Indels 0; Gaps 0;  
QY 9 TGTTCAGTCAACATCTACATAAAAAATAAAAAACAGAAAAAGAAATGAAGAAAGC 68  
Db 616 TATCTCAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 557  
QY 69 CTGAGCCAGATATTGGTATTAAGAGGAGCTGAAGAAAAAGAGAGACAAAGAGAGG 128  
Db 556 AAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 497  
QY 129 AAAAAAGGAAATTTGAAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 188  
Db 496 AAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 437  
QY 189 AAAAAATATTAAG 248  
Db 436 AAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 377  
QY 249 AGTGAAGGAGAGAA 264  
Db 376 AAAAAAATAAAAAA 361  
RESULT 11  
ABL33404/c  
ID ABL33404 standard; DNA; 16033 BP.  
XX ABL33404;  
XX 26-MAR-2002 (first entry)  
XX Human immune system associated gene SEQ ID NO: 1377.  
XX

KW Human; immune system disease; cytosine methylation; antiasthmatic;  
KW antiarteriosclerotic; antianaemic; cytostatic; neurotropic;  
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;  
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;  
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;  
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;  
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;  
KW gene; ds.  
XX Homo sapiens.  
XX WO200200928-A2.  
XX 03-JAN-2002.  
XX 02-JUL-2001; 2001WO-EP07537.  
XX 30-JUN-2000; 2000DE-1032529.  
XX 01-SEP-2000; 2000DE-1043826.  
XX (EPIG-) EPIGENOMICS AG.  
XX Olek A, Piepenbrock C, Berlin K;  
XX WPI; 2002-130909/17.  
XX Nucleic acid comprising fragment of chemically modified gene, useful  
PT for diagnosis and treatment of diseases associated with abnormal  
PT cytosine methylation  
XX Claim 1; SEQ ID NO 1377; 32pp + Sequence Listing; German.  
XX The present invention provides a number of human immune system associated  
CC genes which are modified by the methylation of cytosines. The sequences  
CC can be used in the diagnosis and treatment of immune system disorders,  
CC including eye diseases such as retinopathy, neovascular glaucoma and  
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
CC diseases. The present sequence is a gene of the invention.  
XX Sequence 16033 BP; 3610 A; 352 C; 4561 G; 7510 T; 0 other;  
Query Match 37.9%; Score 100.4; DB 24; Length 16033;  
Best Local Similarity 63.0%; Pred. No. 2.4e-07;  
Matches 155; Conservative 0; Mismatches 91; Indels 0; Gaps 0;  
QY 19 AACATCTACATATAAAAAATAAAAAACAGAAAAAGAAATGAAGAAAGCTGAGCCAGA 78  
Db 3372 AAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 3313  
QY 79 TATTCGTATTAAG 138  
Db 3312 AAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 3253  
QY 139 GGAATTTGAAGATTA 198  
Db 3252 AAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 3193  
QY 199 TAAAGAAAG 258  
Db 3192 TAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 3133  
QY 259 AGAGAA 264  
Db 3132 AAAAAA 3127  
RESULT 12  
ABL33719/c  
ID ABL33719 standard; DNA; 17934 BP.  
XX

```
AC ABL33719;
XX
XX 26-MAR-2002 (first entry)
DE Human immune system associated gene SEQ ID NO: 1692.
DE
XX
XX Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytosatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW gene; ds.
XX
XX Homo sapiens.
OS
XX WO200200928-A2.
XX
XX 03-JAN-2002.
XX
XX 02-JUL-2001; 2001WO-EP07537.
XX
XX 30-JUN-2000; 2000DE-1032529.
XX
XX 01-SEP-2000; 2000DE-1043826.
XX
XX (EPIG-) EPIGENOMICS AG.
XX
XX Olek A, Piepenbrock C, Berlin K;
XX WPI; 2002-130909/17.
XX
XX Nucleic acid comprising fragment of chemically modified gene, useful
XX for diagnosis and treatment of diseases associated with abnormal
XX cytosine methylation -
XX
XX Claim 1; SEQ ID NO 1692; 32pp + Sequence Listing; German.
XX
XX The present invention provides a number of human immune system associated
XX genes which are modified by the methylation of cytosines. The sequences
XX can be used in the diagnosis and treatment of immune system disorders,
XX including eye diseases such as retinopathy, neovascular glaucoma and
XX macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
XX leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
XX rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
XX diseases. The present sequence is a gene of the invention.
XX
XX Sequence 17934 BP; 4426 A; 283 C; 4402 G; 8823 T; 0 other;

Query Match 37.9%; Score 100.4; DB 24; Length 17934;
Best Local Similarity 63.0%; Pred. No. 2.4e-07;
Matches 155; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

QY 19 AACATCTACATAAAAAATAAACAAGAAAAAGAAAAATGAAGAAAGCCTGAGCCAGA 78
D 17855 AAAAAAAAAAAAAAAAAATCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 17796
QY 79 TATTGGTATAAAGAGGAGCTGAAGAAAAAGAGAGACAAAGAGAGAGAGAAATAAAG 138
D 17795 AAAAAAAAAAAAAAAAAAACTAAAAAATAAAAAAAAAAAAAAAAAAAAAA 17736
QY 139 GGAATTGAAAGGAGAAAAAGAAAAAGAGGATAAAGAAAGAAATTAAGAAAAAGATA 198
D 17735 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 17676
QY 199 TAAAGAAAGAGAGAAAAACAAGTAAAGAAATCCACAGAGAAAGAAAGAAAGTGAAGGA 258
D 17675 AAAAAAAAAAAAAAAAAATAAAAAAAAAAAAAAAAAAAAAAACAATAAATAA 17616
QY 259 AGAGAA 264
D 17615 AAAAAA 17610
```

```
RESULT 13
ABL33958/C
ID ABL33958 standard; DNA; 14006 BP.
XX
XX ABL33958;
AC
XX 26-MAR-2002 (first entry)
DT
XX Human immune system associated gene SEQ ID NO: 1931.
DE
XX Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytosatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW gene; ds.
XX
XX Homo sapiens.
OS
XX WO200200928-A2.
XX
XX 03-JAN-2002.
XX
XX 02-JUL-2001; 2001WO-EP07537.
XX
XX 30-JUN-2000; 2000DE-1032529.
XX
XX 01-SEP-2000; 2000DE-1043826.
XX
XX (EPIG-) EPIGENOMICS AG.
XX
XX Olek A, Piepenbrock C, Berlin K;
XX WPI; 2002-130909/17.
XX
XX Nucleic acid comprising fragment of chemically modified gene, useful
XX for diagnosis and treatment of diseases associated with abnormal
XX cytosine methylation -
XX
XX Claim 1; SEQ ID NO 1931; 32pp + Sequence Listing; German.
XX
XX The present invention provides a number of human immune system associated
XX genes which are modified by the methylation of cytosines. The sequences
XX can be used in the diagnosis and treatment of immune system disorders,
XX including eye diseases such as retinopathy, neovascular glaucoma and
XX macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
XX leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
XX rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
XX diseases. The present sequence is a gene of the invention.
XX
XX Sequence 14006 BP; 3278 A; 155 C; 3257 G; 7313 T; 3 other;

Query Match 37.7%; Score 99.8; DB 24; Length 14006;
Best Local Similarity 62.8%; Pred. No. 3e-07;
Matches 155; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

QY 18 GAAACATCTACATAAAAAATAAACAAGAAAAAGAAATGAAGAAAGCCTGAGCCAG 77
D 2301 GAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2242
QY 78 ATATTGGTATAAAGAGGAGCTGAAGAAAAAGAGAGACAAAGAGAGAAATAAAA 137
D 2241 AAAAAAAAAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2182
QY 138 GCGAATTGAAAGGAGAGAAAAAGAAAAAGAGGATTAAGAAAGAAATTAAGAAAAAGATA 197
D 2181 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATCAAAAAAAAAAAAAAAAAA 2122
QY 198 ATAAAGAAAGAGAGAAAAACAAAGTAAAGATCCACACAGAAAGAAAGAGTGAAGG 257
```

80	ATTTCGTTATTAACAGGAGCTGAAGAAAGAAAGAGAGAGCAAAAGAGAGAGGAAGAAATTAAGG	139
Qy		
269	AA	210
Db		
140	GAATTGAAAGGGAGAAAGAAAGAGAGATAGAAAGAAATTTAAAGAGAAAAAGATTAAT	199
Qy		
209	AA	150
Db		
200	AAAGAAAGAGAGAAACCAAGTAAAGAAATCCACACAGAAAGAAAAAGAGTGAAGCAA	259
Qy		
149	AA	90
Db		

200	AAAGAAAAAGAGAAAAACAAAGTAAAGAAATCCACACAGAAAGAAAAAGAGTGAAGAA	259
QY	111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111	
Db	149 AAAAAA	90
	260 GAGAA 264	
QY	89 AAAAA 85	
Db		

Db 89 AAAAA 85

RESULT 15  
AAS63367/C  
ID AAS63367 standard; DNA; 3586 BP.  
XX  
AC AAS63367;

XX  
DT 29-JAN-2002 (first entry)

Chemically pretreated metabolism associated gene #62.

XX Human; cytostatic; anti-tumour; metabolism; metabolic disease; liver;  
KW

KW solid tumour; cancer; cytosine methylation; epigenetic; eye; kidney;  
 KW single nucleotide polymorphism detection; SNP; stool; urine; lung;

KW cerebral-spinal fluid; intestine; brain; heart; prostate; breast;  
 KW DUSP2; EPHX2; QDPR; SGSH; SHMT2; SLC7A2; SLC7A4; TYMS; ds.  
 KW

xx  
OS Homo sapiens.

XX  
PN  
W0300176451-A2.

XX  
PD  
18-OCT-2001.

XX  
PF  
06-APR-2001: 2001WO-EP04016.XX  
PR 06-APR-2000: 2000DE-1019058.

PR 07-APR-2000; 2000DE-1019173.  
PR 30-JUN-2000; 2000DE-1032529.

PR 01-SEP-2000; 2000DE-1043826.  
XX  
XX

PA (EPIG-) EPIGENOMICS AG.

PI  
XX  
olek A, Piepenbrock C, Berlin K;

WPI; 2002-010834/01.

New nucleic acid, useful for diagnosis and therapy of metabolic disease solid tumour and cancers comprises segment of chemically

modified genomic sequences of genes associated with metabolism -

Claim 1; Page 141-142; 143pp; English.

The invention relates to a nucleic acid (I) comprising a sequence at

CC associated with metabolism such as DUSP2 (NM\_004418), EPHX2 (NM\_00197  
CC least 10 bases of a segment of the chemically preclated DNA of gene  
CC

CC (NM\_000320), SGGH (NM\_000159), SHH2 (NM\_003412), SLC7A2  
CC (NM\_003046), SLC7A4 (NM\_004173) and TYMS (NM\_001071) (all  
CC undefined). T1 are useful for diagnosis and therapy of metabolic

underlined). (1) are useful for diagnosis and therapy of metabolic disease, solid tumours and cancers; as primer oligonucleotides for the amplification of DNA sequences for detecting the extent of methylation

state and/or single nucleotide polymorphisms (SNPs) in a chemically treated DNA of cancer associated with methylation in cancer (7).

CC treated DNA of genes associated with metabolism. An array of (1) is  
CC useful for ascertaining genetic and/or epigenetic parameters for the

CC diagnosis and/or therapy of existing diseases or the predisposition  
CC specific diseases by analysing cytosine methylations. The method involves

CC chemically treating genomic DNA sample by a solution of bisulphite,  
CC hydrogen sulphite or disulphite such that cytosine bases which are



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 2, 2002, 16:53:13 ; Search time 8940.62 Seconds  
(without alignments)  
400.050 Million cell updates/sec

Title: US-08-973-363-12

Perfect score: 265

Sequence: 1 GATGAGATTGTTTCAGTGAA.....AAGAAGTGAAGGAGAGAG 265

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_estl:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_gss:\*  
13: em\_gss\_hum:\*  
14: em\_gss\_inv:\*  
15: em\_gss\_pin:\*  
16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	178	67.2	677	10	BM015597
2	151.6	57.2	821	10	BF239967
3	143.8	54.3	430	9	A1242163
4	142.2	53.7	343	9	AA748563
5	139.8	52.8	455	9	AA699318
6	107.8	40.7	851	3	BI644529
7	107.8	40.7	982	10	BM416142
8	106.8	40.3	697	12	AG047727
9	106.4	40.2	812	12	AG044252
10	106.4	40.2	820	12	AG058657
11	105.6	39.3	1101	12	CNS0160P
12	105.6	39.8	338	12	CNS006AG
13	105.4	39.8	748	12	AG031854
14	105.2	39.7	357	12	CNS009YU
15	105.2	39.7	749	12	AG040837
16	105.2	39.7	978	12	CNS06PJK
17	105	39.6	888	12	AZ549422

c 18	104.8	39.5	614	12	AQ327327
19	104.8	39.5	647	12	AG044222
20	104.8	39.5	796	12	AG063366
c 21	104.8	39.5	838	12	CNS03W9J
c 22	104.8	39.5	971	10	BM414980
c 23	104.6	39.5	388	12	CNS021KE
c 24	104.6	39.5	898	12	AZ540362
c 25	104.6	39.5	961	12	AG072942
c 26	104.6	39.5	964	12	AZ667479
c 27	104.6	39.5	982	12	AQ325799
c 28	104.4	39.4	335	12	CNS01TKI
c 29	104.4	39.4	817	12	CNS009FM
30	104.4	39.4	887	12	AZ175011
31	104.4	39.4	963	12	CNS00A4L
32	104.2	39.3	313	12	CNS00C2T
33	104.2	39.3	357	12	CNS00C2V
c 34	104.2	39.3	700	9	AI906328
35	104.2	39.3	952	12	AQ897460
36	104	39.2	849	12	AG044091
c 37	104	39.2	937	12	CNS0066L
c 38	104	39.2	952	12	CNS014BF
c 39	103.8	39.2	431	12	CNS00YW2
c 40	103.8	39.2	830	3	BI645072
c 41	103.8	39.2	966	10	BM415686
c 42	103.8	39.2	990	12	CNS006OI
c 43	103.6	39.1	723	12	AG031890
c 44	103.6	39.1	868	12	AZ544118
c 45	103.6	39.1	979	12	AZ538989

#### ALIGNMENTS

RESULT 1  
BM015597  
LOCUS 603641828F1 NIH\_MGC\_87 Homo sapiens cDNA clone IMAGE:5417887 5',  
DEFINITION mRNA sequence.  
ACCESSION BM015597.1 GI:16529951  
VERSION BM015597.1  
KEYWORDS EST  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 677)  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs@mail.nih.gov  
Tissue Procurement: DCTD/DTF

cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: L1AM12066 row: 9 column: 08  
High quality sequence stop: 674.

#### FEATURES

Location/Qualifiers  
1..677  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5417887"  
/clone\_lib="NIH\_MGC\_87"  
/tissue\_type="mammary adenocarcinoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: breast; Vector: pCMV-SPORT6; Site: 1. Noti; Site: 2. Sall; Cloned unidirectionally; oligo-dT primed. Average insert size 1.383 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH\_MGC Library."

BASE COUNT 305 a 92 c 153 g 127 t  
ORIGIN

Query Match 67.2%; Score 178; DB 10; Length 677;  
Best Local Similarity 84.2%; Pred. No. le-09;  
Matches 224; Conservative 0; Mismatches 40; Indels 2; Gaps 2;  
  
QY 1 GATGAGATTGTTTCAGTGAACATCTACATAAAAAATAAAAAACAGAAAAAAGAAATGA 60  
DB 90 GATGAGATCAGTCTGTGAAACATCCCAATATAAAAAATTAACACAGAAAG-AGACAGTGA 148  
QY 61 AGAAAAGCCTGAGCCAGATATTGTTATATAAGAGAAAGCTGAAGAAAAAGAGAGACAAA 120  
DB 149 AGAAAAGCCTGAGCCAGATGTTTATATAAGAGAAAGCAACCAAGAGAGGGAAGCAA 208  
QY 121 AGAAGAGAAATTAAGAGGGAATTTGAAGAGGAGAAAAAGAAAGAGAGATTAAGAA 180  
DB 209 AGAAGAGGAGAAATAAGAGAACTTAAAGGGGAGATATAAGAGAAAGAGGATTAAGAA 268  
QY 181 ATTAAGAGAAAAAGATATAAGAAAAAGAGAAAAACAAAGTAAAAAGATCCACACAGAA 240  
DB 269 TATAAGAGAAAGATTTTAAGAAAAAGAGAAACAAAGATAAGAAAGCTATACAGAA 328  
QY 241 -AGAAAAGAGTGAAGGAGAGAG 265  
DB 329 GAGAAAAAGACATAAGAGGAGAAAG 354

## RESULT 2

BF239967 601905170F1 NIH\_MGC\_54 821 bp mRNA linear EST 14-NOV-2000  
LOCUS  
DEFINITION mRNA sequence.

ACCESSION BF239967.1 GI:11153890  
VERSION BF239967.1  
KEYWORDS EST.  
SOURCE human.

## ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 821)

NIH-MGC <http://mgc.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: [cga@r-mail.nih.gov](mailto:cga@r-mail.nih.gov)

Tissue Procurement: AFCC

CDNA Library Preparation: CLONTECH Laboratories, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: L1CMI033 row: k column: 18

High quality sequence stop: 562.

Location/Qualifiers

## FEATURES

source

1. .821  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4133129"  
/tissue\_type="from Chronic myelogenous leukemia"  
/lab\_host="DH10B (T1 phage-resistant)"  
/note="Organ: bone marrow; Vector: pDNR-LIB (Clontech);  
Site\_1: sfii (ggcccttcggcc); Site\_2: sfii (ggccattatggcc  
); Double-stranded cDNA was prepared from cell line RNA.  
5' and 3' adaptors were used in cloning as follows: 5'  
adaptor sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor  
sequence: 5'-ATTCTAGAGCGCGAGCGCCGACATG-dt(30)BN-3'  
(where B = A, C, or G and N = A, C, G, or T). Average  
insert size 1.75 kb (range 0.9-4.0 kb). 15/15 colonies  
contained inserts by PCR. This library was enriched for

full-length clones and was constructed by Clontech  
Laboratories (Palo Alto, CA).  
BASE COUNT 345 a 149 c 190 g 137 t  
ORIGIN

Query Match 57.2%; Score 151.6; DB 10; Length 821;  
Best Local Similarity 78.0%; Pred. No. 3.8e-07;  
Matches 195; Conservative 0; Mismatches 54; Indels 1; Gaps 1;

QY 1 GATGAGATTGTTTCAGTGAACATCTACATAAAAAATAAAAAACAGAAAAAAGAAATGA 60

DB 234 GATGAGATCAGTCTGTGAAACATCCCAATATAAAAAATTAACACAGAAAGAGCAGTGA 293

QY 61 AGAAAAGCCTGAGCCAGATATTGTTATATAAGAGAAAGCTGAAGAAAAAGAGAGACAAA 120

DB 294 AGAAAAGCCTGAGCCAGATGTTTATATAAGAGAAAGCAACCAAGAGAAAGAGGGAAGCAA 353

QY 121 AGAAGAGAAATTAAGAGGGAATTTGAAGAGGAGAAAAAGAGAGAGATTAAGAAAG 179

DB 354 AGAAAAGAGATTAAGAGAAAGCACTTAAGAGTGAGATTAAGAAAAAGAGAGATTAAGAAAG 413

QY 180 ATTTAAAGAAAAAGATTAATAAGAAAGAGAGAAAAACAAAGTAAAAAGAAATCCACACAGA 239

DB 414 ATATAAGGAACAAGATTTTAAAGAAAAACGAGAAAAACAAAGTAAAGAGAGCTATACAGA 473

QY 240 AAGAAAAAGA 249

DB 474 ACGAAAAAGA 483

## RESULT 3

AI242163/c

LOCUS

DEFINITION

AI242163 430 bp mRNA linear EST 01-DEC-1998  
qh81f02.x1 Soares\_fetal\_liver\_spleen\_INFLS\_S1 Homo sapiens cDNA  
clone IMAGE:1851099 3' similar to contains element LTR5 repetitive  
element ;, mRNA sequence.

ACCESSION AI242163.1 GI:3837560

VERSION AI242163.1

KEYWORDS EST.

SOURCE human.

## ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 430)

NCI-CCGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: [cga@r-mail.nih.gov](mailto:cga@r-mail.nih.gov)

This clone is available royalty-free through LLNL; contact the

IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.

Insert Length: 1274 Std Error: 0.00

Seq primer: -400P from Gibco

High quality sequence stop: 429.

Location/Qualifiers

## FEATURES

source

1. .430  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1851099"  
/clone\_lib="Soares\_fetal\_liver\_spleen\_INFLS\_S1"  
/sex="male"  
/dev\_stage="20 week-post conception fetus"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)  
with a modified polylinker; Site\_1: Pac I; Site\_2: Eco RI;  
This is a subtracted version of the original Soares fetal  
liver spleen INFLS library. 1st strand cDNA was primed  
with a Pac I oligo(dT) primer [5'  
AACTGGAAGAATTAATAAGATCTTTTTTTTTTTTTTTTTTTT 3']  
double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Pac I and cloned into the Pac I





/note="Organ: Liver and Spleen; Vector: pT73b (Pharmacia) with a modified polylinker; Site.1: Pac I; Site.2: Eco RI; This is a subtracted version of the original Soares fetal liver spleen INFLS library. 1st strand cDNA was primed with a Pac I - oligo(dT) primer [5'-AACTGGAAGATTAATAAGATCTTTTTTTTTTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 119 a 95 c 57 g 185 t  
ORIGIN

Query Match 52.8%; Score 139.8; DB 9; Length 456;  
Best Local Similarity 83.7%; Pred. No. 7.6e-06;  
Matches 170; Conservative 0; Mismatches 32; Indels 1; Gaps 1;

QY 1 GATGAGATTGTTTCAGTGAACATCTACATATAAAAAATAAAAAACAGAAAAAGAAAAATGA 60  
||||| ||| ||||||| ||| ||||||| ||| ||||||| ||| |||  
Db 202 GATGAGATCAGTTCTGTGAACATCCAAATAAAAAATTAAACACAGAAAG-AGACAGTGA 144  
QY 61 AGAAAAGCCGTGACCCAGATATTGGTATAAGAAGGAGCTGAAGAAAAAGAGACACAA 120  
||||| ||||||| ||| ||||||| ||| ||||||| ||| |||  
Db 143 AGAAAACCTGACCCAGATGTTTATATAAGAAGGAGCAACAGAAAGAGGGAAGCAAA 84  
QY 121 AGAAGAGGAAAAATAAAGGAATTAAGGGGAGAAAAAGAAAAAGAGGATTAAGAAGA 180  
||||| ||||||| ||| ||||||| ||| ||||||| ||| |||  
Db 83 AGAAAAGGAGAAATAAAAAAGAACTTAAGGGGAGATAAAAAAGAAAAAGAGGATTAAGAAGA 24  
QY 181 ATTAAGAAAAAGATAATAAG 203  
||||| ||||||| |||  
Db 23 TATAAGGAAAAAGATTTTAAG 1

RESULT 6  
BI644529/c  
ID BI644529 standard; RNA; EST; 851 BP.  
XX BI644529;  
XX BI644529.1  
SV 13-SEP-2001 (Rel. 69, Created)  
DT 13-SEP-2001 (Rel. 69, Last updated, Version 1)  
XX  
DE OP2269 Mixed Stage EST's from Globodera pallida, the potato cyst nematode  
DE Globodera pallida cDNA, mRNA sequence.  
XX  
KW EST.  
XX  
OS Globodera pallida  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;  
OC Tylenchoidea; Heteroderidae; Heteroderinae; Globodera.  
XX  
RN [1]  
RP 1-851  
RA Heer J., Sosinski B., Pokrzywa R.M., Warry A., Opperman C.;  
RT "Mixed Stage EST's from Globodera pallida, the potato cyst nematode";  
RL Unpublished.  
XX  
CC Contact: Opperman, C  
CC Center for the Biology of Nematode Parasitism  
CC NC State University; IACR-Rothamsted  
CC Campus Box 7616; Raleigh, NC 27695, USA  
CC Tel: 919.515.6699  
CC Fax: 919.515.9500  
CC Email: warthog@unity.ncsu.edu  
CC No homology found. ; GT11PCNL\_G06\_1-42F\_040.ab1.seq.screen.  
XX  
FH Key Location/Qualifiers

FT source 1..851  
FT /db\_xref="taxon:36090"  
FT /note="Vector: lambda GT11: This is a collaborative effort  
FT between IACR-Rothamsted and North Carolina State  
FT University. The library was constructed from mixed stage G.  
FT pallida in lambda GT11 by Paul Burroughs,  
FT IACR-Rothamsted."  
FT /organism="Globodera pallida"  
FT /clone\_lib="Mixed Stage EST's from Globodera pallida, the  
FT potato cyst nematode"  
XX  
SQ Sequence 851 BP; 114 A; 72 C; 15 G; 615 T; 35 other;

Query Match 40.7%; Score 107.8; DB 3; Length 851;  
Best Local Similarity 64.8%; Pred. No. 0.0083;  
Matches 160; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

QY 18 GAAACATCTACATAAAAAATAAAAAACAGAAAAAGAAAAATGAAGAAAGCCGTGAGCCAG 77  
||||| ||| ||||||| ||| ||||||| ||| ||||||| |||  
Db 821 GAAAAAAGAAAAAAGAAAAAAGAAAAAAGTAAACAGAAAAAGAAAGGAAA 762  
QY 78 ATATTGCTATAAGAGAGGAGCTGAAGAAAAAGAGACACAAAAAGAGAGGAAAAATAAAA 137  
||||| ||| ||||||| ||| ||||||| ||| ||||||| |||  
Db 761 AAAAAATAAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA 702  
QY 138 GCGAATTGAAAGGGAGAAAAAGAAAAAGAGGATTAAGAAAGATTAAGAAAGAAAAAGATA 197  
||||| ||| ||||||| ||| ||||||| ||| ||||||| |||  
Db 701 TATAATAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA 642  
QY 198 ATAAAGAAAAAGAGAAAAACAAAGTAAAGAAATCCACACAGAAAAAGAAAGAGTGAAGG 257  
||||| ||| ||||||| ||| ||||||| ||| ||||||| |||  
Db 641 AGAAGAAAAAGAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAT 582  
QY 258 AGAGAA 264  
||| |||  
Db 581 AAAAAA 575

RESULT 7  
BM416142/c  
LOCUS  
DEFINITION  
OP21229 Mixed Stage EST's from Globodera pallida, the potato cyst  
nematode Globodera pallida cDNA, mRNA sequence.  
ACCESSION BM416142  
VERSION BM416142.1 GI:18382942  
KEYWORDS EST.  
SOURCE  
ORGANISM  
Globodera pallida.  
Globodera pallida  
Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;  
Tylenchoidea; Heteroderidae; Heteroderinae; Globodera.  
REFERENCE  
1 (bases 1 to 982)  
AUTHORS Heer,J., Sosinski,B., Pokrzywa,R.M., Warry,A. and Opperman,C.  
TITLE Mixed Stage EST's from Globodera pallida, the potato cyst nematode  
JOURNAL Unpublished (2001)  
COMMENT Contact: Opperman, C  
Center for the Biology of Nematode Parasitism  
NC State University; IACR-Rothamsted  
Campus Box 7616; Raleigh, NC 27695, USA  
Tel: 919.515.6699  
Fax: 919.515.9500  
Email: warthog@unity.ncsu.edu  
GT11PCNL\_G06\_1-42F\_040.ab1.  
FEATURES  
source  
1..982  
Location/Qualifiers  
/organism="Globodera pallida"  
/db\_xref="taxon:36090"  
/clone\_lib="Mixed Stage EST's from Globodera pallida, the  
potato cyst nematode"  
/note="Vector: lambda GT11: This is a collaborative effort  
between IACR-Rothamsted and North Carolina State  
University. The library was constructed from mixed stage  
G. pallida in lambda GT11 by Paul Burroughs,



Query Match 40.2%; Score 106.4; DB 12; Length 820;  
Best Local Similarity 52.9%; Pred. No. 0.012;  
Matches 158; Conservative 0; Mismatches .93; Indels 0; Caps 0;

QY	14	CAGTGAACATCTACATAAAAAAATATAAAACAGAAAAAAGAATGAAAGAAAGCCCTGGAG	73
Db	230	CCGTAAATAANTAAAGAAAAAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	289
QY	74	CCGATATTTCGTATAAAGAGGAGCTGAAGAAAAAGAGACACAAAAAGAGAGGAAAAAT	133
Db	290	AAAAAANAANNAA	349
QY	134	AAAGGGAATTGAAAGGGAGAAAAAGAAAAAGAGGATTAAGAAAGAATTAAAGAAAAA	193
Db	350	AAAAAAAAGAAAAAGAAAAAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	409
QY	194	GATAATAACAAGAGAGAAAAACAAAGTAAAGAAATCCACACAGAAAGAAAAAGAGAGTG	253
Db	410	AAAAAAAAAAAAAAAAAAAAAAAAATAAAATAAAAAAAAAAAAAAAAAAAAAAAAAA	469
QY	254	AAGGAAGAGAA	264
Db	470	AAAAAGAAAA	480

RESULT	11	CNS0160P	1101 bp	DNA	linear	GSS 26-JUL-1999
LOCUS		CNS0160P				
DEFINITION		Drosophila melanogaster genome survey sequence T7 end of BAC BACN16H02 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.				
ACCESSION		AL107011				
VERSION		AL107011.1	GI:5624989			
KEYWORDS		GSS.				
SOURCE		fruit fly.				
ORGANISM		Drosophila melanogaster Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.				
REFERENCE		1	(bases 1 to 1101)			
AUTHORS		Genoscope.				
TITLE		Direct Submission				
JOURNAL		Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage ; BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - web : <a href="http://www.genoscope.cns.fr">www.genoscope.cns.fr</a> )				
COMMENT		Determination of this BAC-end sequence was carried out as part of a				

**FEATURES**  
Source  
1. .1101  
Location/Qualifiers  
pBelobAC11.  
and Genevieve Payan. It has been constructed in the vector  
project grant. The DNA was prepared from embryos by Alain Bucheton  
d'Etude du Polymorphisme Humain) with funding provided by a MRC  
library (Dros BAC) was made by Alain Billaud at CEPH (Centre  
http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC  
collaboration with the European Drosophila genome project (EDGP)

```

/organism="BrOsopnopia miranaogaster"
/plasmid="pBelOBAC11"
/db_xref="taxon:7227"
/clone_lib="DrosBAC"
/clone="BACN16H02"
/note="end : T7"

BASE COUNT      549 a    42 c   157 g   227 t   126 others
ORIGIN

Query Match          39.9%; Score 105.8; DB 12; Length 1101;
Best Local Similarity 60.3%; Pred. No. 0.012;
Matches 149; Conservative 13; Mismatches 85; Indels 0; Gaps 0;

QY 19 AAACATCTACATAAAAAAATAAAACAGAAAAGAAAATGAAGAAGGCGCTGAGCCAGA 78
|||||

```

79	QY	TATTGGTATATAAGAGAGAGAGCTGACAGAAAAAAGAGAGACACAAAAGAGAGAGAGAAATAATAAAG	138
274	Db	AAAGARAAAAAARAAAAAARAAAAAARAAAAAARAAAAAARAAAAAARAAAAAARAAAAAARAAAAA	215
139	QY	GGATTGTAAGAGGGAGAAAAAGAAAAAGAGGATTAAAGAAAGAAATTAAAGAAAAAAGATAA	198
214	Db	AAAAAARAAAAAARAAAAAARAAAAAARAAAAAARAAAAAARAAAAAARAAAAAARAAAAA	155
199	QY	TAAAGAAAAAGAGAGAAAAACAAAGCTAAAGAAATCCACACAGAGAAAAAGAAAAAGAGTGAAGGA	258
154	Db	AAAAAAGAGAAAAAARAAAAAARAAAAAARAAAAAARAAAAAARAAAAAARAAAAAARAAAAA	95
259	QY	AGAGAA	264
94	Db	ARARAA	89

RESULT_13	
AG031854	
LOCUS	748 bp DNA linear GSS 01-NOV-2001
DEFINITION	Pan troglodytes DNA, clone: PTB-005017.R, genomic survey sequence.
ACCESSION	AG031854
VERSION	AG031854.1 GI:16558727
KEYWORDS	GSS; GSS (genome survey sequence).
SOURCE	Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
FEATURES	BAC Library clone:PTB-005017.R.

**JOURNAL** Submitted (02-Aug-2001) ASAO Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: chimpbes@gsc.riken.go.jp, URL: <http://hgp.gsc.riken.go.jp/>, Tel: 81-45-503-9111, Fax: 81-45-503-9170)  
**COMMENT** Clones are derived from the chimpanzee BAC library PTB This BAC end was generated during the R6d process and may have higher chance of clone tracking errors.

```

source
1../48
/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone="PTB-005017.R"
/sex="male"
/cell_type="lymphoblast"
/clone.lib="PTB Chimpanzee Male BAC Library"
BASE COUNT 610 a 26 c 76 q 8 t 28 others

```

Query Match	39.8%;	Score 105.4;	DB 12;	Length 748;
Best Local Similarity	62.6%;			
Matches 154;	Conservative	0;	Mismatches 92;	Indels 0;
Gaps 0;				

[illegible]

```
Db 294 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAG 353
QY 139 GGAATTCAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 198
Db 354 GAAAAAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 413
QY 199 TAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 258
Db 414 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 473
QY 259 AGAGAA 264
Db 474 AAAAAA 479

RESULT 14
CNS009YU
LOCUS
DEFINITION
Drosophila melanogaster genome survey sequence T7 end of BAC #
BACR20M04 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION
AL054509
VERSION
AL054509.1 GI:4935682
KEYWORDS
GSS.
SOURCE
fruit fly.
ORGANISM
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
1 (bases 1 to 357)
AUTHORS
Genoscope.
JOURNAL
Direct Submission
TITLE
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osogawa and
Aaron Mamoser in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain v2; cn bw sp, the same strain used for the BDGP's
p1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila\_bac.htm.
FEATURES
source
1. .357
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPCI-98"
/clone="BACR20M04"
/note="end : T7"
BASE COUNT 316 a 9 c 4 g 17 t 11 others
ORIGIN

Query Match 39.7%; Score 105.2; DB 12; Length 357;
Best Local Similarity 63.0%; Pred. No. 0.023;
Matches 155; Conservative 4; Mismatches 87; Indels 0; Gaps 0;

QY 19 AACATCTCATATAAAATATAAAACAGAAAAAGAAATGAAGAAAGCCTGAGCCAGA 78
Db 76 AAAAAAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 135
QY 79 TATTGGTATATAAGAGAGAGCTCAAGAAAAAGAGAGACAAAGAGAGAGAGAGAGAGAG 138
Db 136 AAAAAAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 195
```

```
QY 139 GGAATTCAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 198
Db 196 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 255
QY 199 TAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 258
Db 256 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 315
QY 259 AGAGAA 264
Db 316 AAAAAA 321

RESULT 15
AG040837
LOCUS
DEFINITION
Pan troglodytes DNA, clone: PTB-018F21.F, genomic survey sequence.
ACCESSION
AG040837
VERSION
AG040837.1 GI:16569562
KEYWORDS
GSS: GSS (genome survey sequence).
SOURCE
Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
BAC Library clone:PTB-018F21.F.
ORGANISM
Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE
1 (sites)
AUTHORS
Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE
BAC end sequences of Library PTB
JOURNAL
Unpublished
REFERENCE
2 (bases 1 to 749)
AUTHORS
Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totoki,Y., Watanabe,H. and Sakaki,Y.
JOURNAL
Direct Submission
TITLE
Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou,Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:chimpbes@sc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT
Clones are derived from the chimpanzee BAC library PTB This BAC end
was generated during the R&D process and may have higher chance of
clone tracking errors.
PRIMERS
Sequencing: -21M13
LIBRARY
Vector : PKS145
R.Site 1 : SacI
R.Site 2 : SacI.
FEATURES
Location/Qualifiers
source
1. .749
/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone="PTB-018F21.F"
/sex="male"
/cell_type="lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT 609 a 11 c 80 g 21 t 28 others
ORIGIN

Query Match 39.7%; Score 105.2; DB 12; Length 749;
Best Local Similarity 64.0%; Pred. No. 0.016;
Matches 158; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

QY 19 AACATCTCATATAAAATATAAAACAGAAAAAGAAATGAAGAAAGCCTGAGCCAGA 78
Db 398 AAAAAAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 457
QY 79 TATTGGTATATAAGAGAGAGCTCAAGAAAAAGAGAGACAAAGAGAGAGAGAGAGAGAG 138
Db 458 GAAAGGAAAAAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 517
```

QY 139 GGAATTGAAAGCGAGAAAAAGAAAAAGAGGATAGAAAGAAATTAAAGAAAAAGATTAA 198  
Db 518 AAAAAAAAAAAAAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAA 577  
QY 199 TAAAGAAAGAGAGAAAAACAAAGTAAAGAAATCCACACAGAAAGAAAAAGAAAGTGAAGGA 258  
Db 578 AAAAAAAAAAAAAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAA 637  
QY 259 AGAGAAAG 265  
Db 638 AAAGAAAG 644

Search completed: August 2, 2002, 16:53:17  
Job time: 12278 sec

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 2, 2002, 20:58:08 ; Search time 12674 Seconds  
(without alignments)  
226.207 Million cell updates/sec

Title: US-08-973-363-13  
Perfect score: 137

Sequence: 1 GATGGGATTGTTTCAGTGAA.....AAAAGAAAAGAAATAAGA 137

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:\*

1: gb\_ba:\*

2: gb\_hgt:\*

3: gb\_in:\*

4: gb\_om:\*

5: gb\_ov:\*

6: gb\_pat:\*

7: gb\_ph:\*

8: gb\_pl:\*

9: gb\_pr:\*

10: gb\_ro:\*

11: gb\_sts:\*

12: gb\_sy:\*

13: gb\_un:\*

14: gb\_vi:\*

15: em\_ba:\*

16: em\_fun:\*

17: em\_in:\*

18: em\_mu:\*

19: em\_mu:\*

20: em\_mu:\*

21: em\_or:\*

22: em\_ov:\*

23: em\_pat:\*

24: em\_ph:\*

25: em\_pl:\*

26: em\_ro:\*

27: em\_sts:\*

28: em\_un:\*

29: em\_vi:\*

30: em\_htg\_hum:\*

31: em\_htg\_inv:\*

32: em\_htg\_other:\*

33: em\_htgo\_inv:\*

pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query	ID	Description
------------	-------	-------	-------	----	-------------

1	137	100.0	137	6	A58694
2	121	88.3	265	6	A58693
3	109	79.6	6872	5	AF004397
4	83.4	60.9	101220	9	AC092372
5	83.4	60.9	134365	9	AC012624
6	83.4	60.9	143079	2	AC021449
7	83.4	60.9	145659	2	AC008531
8	83.4	60.9	193446	2	AC009194
9	83.4	60.9	219258	9	AC022121
10	83.4	60.9	276181	2	AC092382
11	57.4	41.9	130030	9	AC004889
12	56.6	41.3	151207	2	AP001031
13	55	40.1	231912	2	AC058806
14	53.8	39.3	152751	2	AC058806
15	53.4	39.0	969	11	CNS06JDL
16	53.4	39.0	23738	2	AC020828
17	53.4	39.0	183099	2	AC025288
18	53.4	39.0	216593	2	AC092126
19	53.2	38.8	42839	2	AC027282
20	53	38.7	7218	6	166494
21	52.6	38.4	3353	6	AX086952
22	52.6	38.4	3353	9	HS0801860
23	52.6	38.4	127580	9	HS077401
24	52.6	38.4	199862	2	AL662809
25	52.6	38.4	215352	2	AL645854
26	52.6	38.4	215352	2	AL645854
27	52.6	38.4	248214	2	AL138811
28	52.4	38.2	165317	9	CNS01RGM
29	52.4	38.2	193354	2	AC025152
30	52	38.0	64593	2	AC099925
31	52	38.0	167069	2	AC067833
32	52	38.0	168922	9	AP002007
33	52	38.0	172859	2	AC010815
34	51.8	37.8	664	9	HS0327335
35	51.8	37.8	173854	2	AC073047
36	51.6	37.7	1018	9	BC022044
37	51.6	37.7	205673	2	AL607044
38	51.4	37.5	63365	2	AC087437
39	51.4	37.5	168463	2	AC021463
40	51.4	37.5	211771	2	AC084821
41	51.4	37.5	213203	2	AC087229
42	51.2	37.4	602	9	BC008488
43	51.2	37.4	1944	10	BC012877
44	51.2	37.4	72874	2	AC087447
45	51.2	37.4	84679	2	AC022571

ALIGNMENTS

RESULT	1	A58694	Sequence 13 from Patent WO9639505.	137 bp	DNA	Linear	PAT 06-MAR-1998
LOCUS	A58694	A58694	GI:3714252				
DEFINITION	A58694	A58694	GI:3714252				
ACCESSION	A58694	A58694	GI:3714252				
VERSION	A58694	A58694	GI:3714252				
KEYWORDS	A58694	A58694	GI:3714252				
SOURCE	A58694	A58694	GI:3714252				
ORGANISM	A58694	A58694	GI:3714252				
REFERENCE	A58694	A58694	GI:3714252				
AUTHORS	A58694	A58694	GI:3714252				
TITLE	A58694	A58694	GI:3714252				
JOURNAL	A58694	A58694	GI:3714252				
COMMENT	A58694	A58694	GI:3714252				
FEATURES	A58694	A58694	GI:3714252				
SOURCE	A58694	A58694	GI:3714252				
BASE COUNT	A58694	A58694	GI:3714252				
ORIGIN	A58694	A58694	GI:3714252				

Patent: WO 9639505-A 13 12-DEC-1996;  
ISIS INNOVATION (GB)  
Other publication AU 5906996 961224.  
Location/Qualifiers  
1. .137  
/organism="unidentified"  
/db\_xref="taxon:32644"  
76 a 10 c 33 g 18 t





REFERENCE  
AUTHORS

2 (bases 1 to 143079)  
Birken, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,  
Anderson, S., Baldwin, J., Barna, N., Beckerly, R., Beda, F.,  
Boguslavsky, L., Bouktal, B., Brown, A., Burkett, G., Castle, A.,  
Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P.,  
DeArelano, K., Dewar, K., Domino, M., Doyle, M., Fenestor, J.,  
Ferreira, P., Fitzhugh, W., Forrest, C., Gage, D., Galagan, J.,  
Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,  
Howland, J.C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,  
Landers, T., Lehoczy, J., Levine, R., Liu, G., Locke, K.,  
Macdonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K.,  
McPheeters, R., Meldrum, J., Meneus, I., Morrow, J., Naylor, J.,  
Norman, C.H., O'Connor, T., O'Donnell, P., Oliver, T.M., Peterson, K.,  
Pierre, N., Pisani, C., Pollara, V., Raymond, C., Riley, R., Rothman, D.,  
Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N.,  
Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,  
Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W.J.,  
Zimmer, A. and Zody, M.

TITLE  
JOURNAL

Direct Submission  
Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA

## COMMENT

On Sep 10, 2000 this sequence version replaced gi:7407963.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

----- Project Information

Center project name: L5154

Center clone name: 58\_M.12

----- Summary Statistics

Sequencing vector: M13; M77815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 134743 bases at least Q40

Consensus quality: 139227 bases at least Q30

Consensus quality: 140814 bases at least Q20

Insert size: 144000; agarose-fp

Insert size: 142179; sum-of-contigs

Quality coverage: 4.6 in Q20 bases; agarose-fp

Quality coverage: 4.7 in Q20 bases; sum-of-contigs

-----

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 10 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 38820: contig of 38820 bp in length  
\* 38821 38920: gap of 100 bp  
\* 38921 40411: contig of 1491 bp in length  
\* 40412 40511: gap of 100 bp  
\* 40512 43279: contig of 2768 bp in length  
\* 43280 43379: gap of 100 bp in length  
\* 43380 46905: contig of 3526 bp in length  
\* 46906 47005: gap of 100 bp  
\* 47006 51830: contig of 4825 bp in length  
\* 51831 51930: gap of 100 bp  
\* 51931 62619: contig of 10689 bp in length  
\* 62620 62719: gap of 100 bp  
\* 62720 75408: contig of 12689 bp in length  
\* 75409 75508: gap of 100 bp  
\* 75509 92516: contig of 17008 bp in length  
\* 92517 92616: gap of 100 bp  
\* 92617 106409: contig of 13793 bp in length  
\* 106410 106509: gap of 100 bp  
\* 106510 143079: contig of 36570 bp in length.

Location/Qualifiers

## FEATURES

## source

1. .143079  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="Rpl1-58M12"  
/clone\_lib="RPC1-11 Human Male BAC"  
1. .38820  
/note="assembly\_fragment"  
clone\_end:SP6  
vector\_side:left"

## misc\_feature

38921..40411  
/note="assembly\_fragment"  
40512..43279  
/note="assembly\_fragment"  
43380..46905  
/note="assembly\_fragment"  
47006..51830  
/note="assembly\_fragment"  
51931..62619  
/note="assembly\_fragment"  
62720..75408  
/note="assembly\_fragment"  
75509..92516  
/note="assembly\_fragment"  
92617..106409  
/note="assembly\_fragment"  
106510..143079  
/note="assembly\_fragment"  
clone\_end:T7  
vector\_side:right"

BASE COUNT 43971 a 26246 c 26678 g 45278 t 906 others  
ORIGIN

Query Match 60.9%; Score 83.4; DB 2; Length 143079;

Best Local Similarity 80.3%; Pred. No. 6e-07;

Matches 110; Conservative 0; Mismatches 26; Indels 1; Gaps 1;

QY 1 GATGGGATGTTTCAGTGAACATCCACATATAAAAAATAAAGCAGAAAAAGAAATGA 60

Db 117669 GATGAGATCAGTCTCTGGAACATCCAAATAAATAATTAACACAGAAAG-AGACAGTGA 117727

QY 61 AGAAAAGATGAGCCAGAGATTGGTATAAGAGAGGAGCTGGAGAAAAAGAGAGACAAA 120

Db 117728 AGAAAACCTGAGCCAGATGTTTATATAAGAGAGACCAAGAAAGAGGGAAGCAA 117787

QY 121 AGAAAAGGAAATAAGA 137

Db 117788 AGAAAAGGAGATATAAA 117804

## RESULT 7

AC008531

LOCUS Homo sapiens chromosome 5 clone CTC-480B11, WORKING DRAFT SEQUENCE,  
DEFINITION 7 ordered pieces.

ACCESSION AC008531

VERSION AC008531.3 GI:12830078

KEYWORDS HTG: HTGS\_PHASE2; HTGS\_DRAFT.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 145659)

AUTHORS DOE Joint Genome Institute.

TITLE Sequencing of Human Chromosome 5

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 145659)

AUTHORS DOE Joint Genome Institute.

TITLE Direct Submission

JOURNAL Submitted (03-Aug-1999) Production Sequencing Facility, DOE Joint

Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

COMMENT On Feb 14, 2001 this sequence version replaced gi:7528342.

-----Genome Center

```
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 369535
Center clone name: CIT-HSPC_480B11
-----
Summary Statistics
Consensus quality: 139128 bases at least Q40
Consensus quality: 142556 bases at least Q30
Consensus quality: 143744 bases at least Q20
Estimated insert size: 148000; pulse field gel estimation
Quality coverage: 6.27 in Q20 bases; sum-of-contigs estimation
Quality coverage: 6.4 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 56174: contig of 56174 bp in length
* 56175 56274: gap of unknown length
* 56275 100874: contig of 44600 bp in length
* 100875 100974: gap of unknown length
* 100975 113127: contig of 12153 bp in length
* 113128 113227: gap of unknown length
* 113228 118190: contig of 4963 bp in length
* 118191 118290: gap of unknown length
* 118291 119694: contig of 1404 bp in length
* 119695 119794: gap of unknown length
* 119795 123297: contig of 3503 bp in length
* 123298 123397: gap of unknown length
* 123398 145659: contig of 22262 bp in length.
FEATURES
Location/Qualifiers
Source
1..145659
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CIT-480B11"
/clone_lib="Caltech human BAC library C"
BASE COUNT 42561 a 26309 c 27580 g 48609 t 600 others
ORIGIN
Query Match 60.9%; Score 83.4; DB 2; Length 145659;
Best Local Similarity 80.3%; Pred. NO. 6e-07;
Matches 110; Conservative 0; Mismatches 26; Indels 1; Gaps 1;
QY 1 GATGGGATTGTTTCAGTGAACATCCACATAAAAAATAAAGCAGAAAAAGAAATGA 60
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 48286 GATGAGATCAGTCTGTGAAACATCCCAATAAAAAATAAAGCAGAAAG-AGACAGTGA 48344
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 61 AGAAAAAGATGACCCAGAGATTGTTAAGAGGAGAGCTGGAGAAAAAGAGACAAA 120
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 48345 AGAAAAACCTGACCCAGATGTTTATATAGAGGAGCCAGAGAAAGAGGAGCAA 48404
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 121 AGAAAGGAAATAGA 137
Db 48405 AGAAAGGAAATAGAA 48421
RESULT 8
AC091946 AC091946 193446 bp DNA linear HTG 09-JUN-2001
LOCUS Homo sapiens chromosome 5 clone RP11-36012, *** SEQUENCING IN
DEFINITION PROGRESS ***, 33 unordered pieces.
ACCESSION AC091946
VERSION AC091946.1 GI:14333882
```

```
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 193446)
DOE Joint Genome Institute.
Sequencing of Human Chromosome 5
Unpublished
JOURNAL
TITLE 2 (bases 1 to 193446)
DOE Joint Genome Institute.
DIRECT SUBMISSION
Submitted (09-JUN-2001) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
JOURNAL
TITLE -----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 544799
Center clone name: RPCI-11_36012
-----
Summary Statistics
Consensus quality: 157767 bases at least Q40
Consensus quality: 180259 bases at least Q30
Consensus quality: 184175 bases at least Q20
Estimated insert size: 204590; agarose-fp estimation
Estimated insert size: 190246; sum-of-contigs estimation
Quality coverage: 5.38 in Q20 bases; agarose-fp estimation
Quality coverage: 5.79 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 33 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1358: contig of 1358 bp in length
* 1359 1458: gap of unknown length
* 1459 2567: contig of 1209 bp in length
* 2568 2767: gap of unknown length
* 2768 3823: contig of 1056 bp in length
* 3824 3823: gap of unknown length
* 3924 5920: contig of 1997 bp in length
* 5921 7494: contig of 1474 bp in length
* 7495 7595: gap of unknown length
* 7596 9096: contig of 1502 bp in length
* 9097 9196: gap of unknown length
* 9197 10260: contig of 1064 bp in length
* 10261 10360: gap of unknown length
* 10361 12460: contig of 2100 bp in length
* 12461 12560: gap of unknown length
* 12561 14511: contig of 2051 bp in length
* 14512 14711: gap of unknown length
* 14712 16381: contig of 1670 bp in length
* 16382 16481: gap of unknown length
* 16482 17968: contig of 1487 bp in length
* 17969 18068: gap of unknown length
* 18069 20434: contig of 2366 bp in length
* 20435 20534: gap of unknown length
* 20535 23515: contig of 2981 bp in length
* 23516 23615: gap of unknown length
* 23616 27563: contig of 3948 bp in length
* 27564 27663: gap of unknown length
* 27664 30987: contig of 3324 bp in length
* 30988 31087: gap of unknown length
* 31088 36065: contig of 4978 bp in length
* 36066 36165: gap of unknown length
* 36166 40978: contig of 4813 bp in length
* 40979 41078: gap of unknown length
```

```
* 41079 45663: contig of 4585 bp in length
* 45664 45763: gap of unknown length
* 45764 51745: contig of 5982 bp in length
* 51746 51845: gap of unknown length
* 51846 57359: contig of 5514 bp in length
* 57360 57459: gap of unknown length
* 57460 67881: contig of 10422 bp in length
* 67882 71132: contig of 6151 bp in length
* 71133 74232: gap of unknown length
* 74233 79795: contig of 5563 bp in length
* 79796 79895: gap of unknown length
* 79896 87511: contig of 7616 bp in length
* 87512 87611: gap of unknown length
* 87612 92791: contig of 5180 bp in length
* 92792 92891: gap of unknown length
* 92892 102794: contig of 9903 bp in length
* 102795 102894: gap of unknown length
* 102895 110866: contig of 7972 bp in length
* 110867 110966: gap of unknown length
* 110967 117571: contig of 6605 bp in length
* 117572 117671: gap of unknown length
* 117672 123738: contig of 6067 bp in length
* 123739 123839: gap of unknown length
* 123839 130584: contig of 6745 bp in length
* 130584 130683: gap of unknown length
* 130684 141544: contig of 10861 bp in length
* 141545 141644: gap of unknown length
* 141645 169109: contig of 27465 bp in length
* 169110 169210: gap of unknown length
* 169210 193446: contig of 24237 bp in length.
FEATURES
    source
        Location/Qualifiers
            1. .193446
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /chromosome="5"
                /clone_lib="RP11-360I2"
                /clone="RP11-360I2"
BASE COUNT 57571 a 35252 c 37061 g 60360 t 3202 others
ORIGIN

Query Match 60.9%; Score 83.4; DB 2; Length 193446;
Best Local Similarity 80.3%; Pred. No. 5.7e-07;
Matches 110; Conservative 0; Mismatches 26; Indels 1; Gaps 1;

QY 1 GATGGGATTGTTTCAGTGAACATCCACATATAAAAAATAAAGCAGAAAAAGAAATGA 60
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 89180 GATGAGATCAGTCTGTGTGAACATCCCAATATAAAAAATAAAGCAGAAAG-AGACAGTGA 89238

QY 61 AGAAAAAGATGAGCCAGAGATTGGTATATAAGAGAGAGCTGGAGAAAAAGAGAGACAAA 120
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 89239 AGAAAAACCTGAGCCAGATGTTTATATAAGAGAGAGACCAAGAGAGAGAGGAGGAGCAA 89298

QY 121 AGAAAAAGAAAAATAAGA 137
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 89299 AGAAAAAGAGAAATAAAA 89315

RESULT 9
AC022121/c 219258 bp DNA linear PRI 30-AUG-2001
LOCUS Homo sapiens chromosome 5 clone CTD-2007H13, complete sequence.
DEFINITION AC022121
ACCESSION AC022121
VERSION AC022121.6 GI:15375145
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 219258)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
```

---

```
Unpublished
2 (bases 1 to 219258)
DOE Joint Genome Institute.
Direct Submission
Submitted (26-JAN-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
3 (bases 1 to 219258)
DOE Joint Genome Institute and Stanford Human Genome Center.
Direct Submission
Submitted (11-AUG-2001) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
4 (bases 1 to 219258)
DOE Joint Genome Institute and Stanford Human Genome Center.
Direct Submission
Submitted (30-AUG-2001) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
On Aug 30, 2001 this sequence version replaced gi:15148108.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40.99.8% of Sequence;
Estimated Total Number of Errors is 0.4.
STS Content:
WI-5811 G04974
WI-13675 G23101
SHGC-58345 G38487
SHGC-103595 G57841.
FEATURES
    source
        Location/Qualifiers
            1. .219258
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /chromosome="5"
                /clone="CTD-2007H13"
BASE COUNT 71954 a 42062 c 40933 g 64309 t
ORIGIN

Query Match 60.9%; Score 83.4; DB 9; Length 219258;
Best Local Similarity 80.3%; Pred. No. 5.6e-07;
Matches 110; Conservative 0; Mismatches 26; Indels 1; Gaps 1;

QY 1 GATGGGATTGTTTCAGTGAACATCCACATATAAAAAATAAAGCAGAAAAAGAAATGA 60
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 87924 GATGAGATCAGTCTGTGTGAACATCCCAATATAAAAAATAAAGCAGAAAG-AGACACTGA 87866

QY 61 AGAAAAAGATGAGCCAGAGATTGGTATATAAGAGAGAGCTGGAGAAAAAGAGAGACAAA 120
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 87865 AGAAAAACCTGAGCCAGATGTTTATATAAGAGAGAGACCAAGAGAGAGGAGGAGCAA 87806

QY 121 AGAAAAAGAAAAATAAGA 137
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 87805 AGAAAAAGAGAAATAAAA 87789

RESULT 10
AC092382/c 276181 bp DNA linear HTG 03-JUL-2001
LOCUS Homo sapiens chromosome 5 clone RP11-75H1, WORKING DRAFT SEQUENCE.
DEFINITION AC092382
ACCESSION AC092382
VERSION AC092382.1 GI:14589571
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 276181)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 5
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 276181)
AUTHORS DOE Joint Genome Institute.
```







149857 151207 contig of 1351 bp in length  
Sequence updated (14-Jan-2000).  
Sequence updated (26-May-2000).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 37 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 15188: contig of 15188 bp in length  
15189 15288: gap of 100 bp  
15289 26305: contig of 11017 bp in length  
26306 26405: gap of 100 bp  
26406 36486: contig of 10081 bp in length  
36487 36586: gap of 100 bp  
36587 44120: contig of 7534 bp in length  
44121 44220: gap of 100 bp  
44221 50920: contig of 6700 bp in length  
50921 51020: gap of 100 bp  
51021 57559: contig of 6539 bp in length  
57560 57659: gap of 100 bp  
57660 65269: contig of 7610 bp in length  
65270 65369: gap of 100 bp  
65370 71743: contig of 6380 bp in length  
71750 71849: gap of 100 bp  
71850 78380: contig of 6331 bp in length  
78381 78480: gap of 100 bp  
78481 83452: contig of 4372 bp in length  
83453 83521: gap of 100 bp  
83553 86778: contig of 3226 bp in length  
86779 86878: gap of 100 bp  
86879 91105: contig of 4227 bp in length  
91106 91205: gap of 100 bp  
91206 95548: contig of 4343 bp in length  
95549 95648: gap of 100 bp  
95649 99866: contig of 4218 bp in length  
99867 99966: gap of 100 bp  
99967 104216: contig of 4250 bp in length  
104217 104316: gap of 100 bp  
104317 107489: contig of 3173 bp in length  
107490 107589: gap of 100 bp  
107590 110861: contig of 3272 bp in length  
110862 110961: gap of 100 bp  
110962 113861: contig of 2900 bp in length  
113862 113961: gap of 100 bp  
113962 115644: contig of 1683 bp in length  
115645 115744: gap of 100 bp  
115745 118658: contig of 2914 bp in length  
118659 118758: gap of 100 bp  
118759 121881: contig of 3123 bp in length  
121882 121981: gap of 100 bp  
121982 124111: contig of 2130 bp in length  
124112 124211: gap of 100 bp  
124212 127985: contig of 3774 bp in length  
127986 128085: gap of 100 bp  
128086 129854: contig of 1769 bp in length  
129855 129954: gap of 100 bp  
129955 132151: contig of 2197 bp in length  
132152 132251: gap of 100 bp  
132252 134285: contig of 2034 bp in length  
134286 134385: gap of 100 bp  
134386 135917: contig of 1532 bp in length  
135918 136017: gap of 100 bp  
136018 137458: contig of 1441 bp in length  
137459 137558: gap of 100 bp  
137559 139352: contig of 1794 bp in length  
139353 139452: gap of 100 bp  
139453 141207: contig of 1755 bp in length  
141208 141307: gap of 100 bp  
141308 142818: contig of 1511 bp in length  
142819 142918: gap of 100 bp

JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Unpublished  
2 (bases 1 to 231912)  
DOE Joint Genome Institute.  
Direct Submission  
Submitted (10-JAN-2001) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
-----Genome Center  
Center: Joint Genome Institute  
Center Code: JGI  
Web site: <http://www.jgi.doe.gov>  
-----  
Project Information  
Center Project Name: 0  
Center clone name: RPCI-23\_127H3  
-----

Summary Statistics  
Consensus quality: 150164 bases at least Q40  
Consensus quality: 175115 bases at least Q30  
Consensus quality: 186307 bases at least Q20  
Estimated insert size: 216300; agarose-fp estimation  
Estimated insert size: 226312; sum-of-contigs estimation  
Quality coverage: 3.5 in Q20 bases; agarose-fp estimation  
Quality coverage: 3.34 in Q20 bases; sum-of-contigs estimation.  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 57 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence.  
\* as soon as it is available and the accession number will  
\* be preserved.

1 1074: contig of 1074 bp in length  
\* 1075 1174: gap of unknown length  
\* 1175 2270: contig of 1096 bp in length  
\* 2271 2370: gap of unknown length  
\* 2371 3511: contig of 1141 bp in length  
\* 3512 3611: gap of unknown length  
\* 3612 4783: contig of 1172 bp in length  
\* 4784 4883: gap of unknown length  
\* 4884 6377: contig of 1494 bp in length  
\* 6378 6477: gap of unknown length  
\* 6478 7111: contig of 1234 bp in length  
\* 7112 7812: gap of unknown length  
\* 7812 9093: contig of 1281 bp in length  
\* 9093 10330: gap of unknown length  
\* 10331 10430: contig of 1138 bp in length  
\* 10431 11677: contig of 1247 bp in length  
\* 11678 11777: gap of unknown length  
\* 11778 13186: contig of 1409 bp in length  
\* 13187 13286: gap of unknown length  
\* 13287 14565: contig of 1279 bp in length  
\* 14566 14665: gap of unknown length  
\* 14666 15787: contig of 1122 bp in length  
\* 15788 15887: gap of unknown length  
\* 15888 17012: contig of 1125 bp in length  
\* 17013 17112: gap of unknown length  
\* 17113 18353: contig of 1241 bp in length  
\* 18354 18453: gap of unknown length  
\* 18454 19611: contig of 1158 bp in length  
\* 19612 19711: gap of unknown length  
\* 19712 21084: contig of 1373 bp in length  
\* 21085 21184: gap of unknown length  
\* 21185 22398: contig of 1214 bp in length  
\* 22399 22498: gap of unknown length  
\* 22499 23629: contig of 1131 bp in length  
\* 23630 23729: gap of unknown length  
\* 23730 25112: contig of 1383 bp in length  
\* 25113 25212: gap of unknown length  
\* 25213 26529: contig of 1317 bp in length  
\* 26530 26629: gap of unknown length  
\* 26630 27773: contig of 1144 bp in length  
\* 27774 27773: gap of unknown length

27874 29211: contig of 1338 bp in length  
29212 29311: gap of unknown length  
29312 30438: contig of 1127 bp in length  
30439 30538: gap of unknown length  
30539 32030: contig of 1392 bp in length  
31931 32031: gap of unknown length  
32031 33112: contig of 1082 bp in length  
33113 33212: gap of unknown length  
33213 34380: contig of 1168 bp in length  
34381 34480: gap of unknown length  
34481 35515: contig of 1035 bp in length  
35516 35615: gap of unknown length  
35616 37145: contig of 1530 bp in length  
37146 37245: gap of unknown length  
37246 38255: contig of 1410 bp in length  
38256 38755: gap of unknown length  
38756 40418: contig of 1663 bp in length  
40419 40518: gap of unknown length  
40519 42089: contig of 1571 bp in length  
42090 42189: gap of unknown length  
42190 43485: contig of 1296 bp in length  
43486 43585: gap of unknown length  
43586 45557: contig of 1972 bp in length  
45558 45657: gap of unknown length  
45658 47561: contig of 1904 bp in length  
47562 47661: gap of unknown length  
47662 50270: contig of 2609 bp in length  
50271 50370: gap of unknown length  
50371 52463: contig of 2093 bp in length  
52464 52563: gap of unknown length  
52564 54241: contig of 1678 bp in length  
54242 54341: gap of unknown length  
54342 56282: contig of 1941 bp in length  
56283 56382: gap of unknown length  
56383 57724: contig of 1342 bp in length  
57725 57824: gap of unknown length  
57825 60404: contig of 2580 bp in length  
60405 60504: gap of unknown length  
60505 63040: contig of 2536 bp in length  
63041 63140: gap of unknown length  
63141 64314: contig of 1174 bp in length  
64315 64414: gap of unknown length  
64415 68063: contig of 3649 bp in length  
68064 68163: gap of unknown length  
68164 72715: contig of 4552 bp in length  
72716 72815: gap of unknown length  
72816 79858: contig of 7043 bp in length  
79859 79958: gap of unknown length  
79959 87305: contig of 7347 bp in length  
87306 87405: gap of unknown length  
87406 94304: contig of 6899 bp in length  
94305 94404: gap of unknown length  
94405 99973: contig of 5569 bp in length  
99974 100073: gap of unknown length  
100074 110741: contig of 10668 bp in length  
110742 110841: gap of unknown length  
110842 116556: contig of 5715 bp in length  
116557 116656: gap of unknown length  
116657 126137: contig of 9481 bp in length  
126138 126237: gap of unknown length  
126238 139661: contig of 13424 bp in length  
139662 139761: gap of unknown length  
139762 147923: contig of 8162 bp in length  
147924 148023: gap of unknown length  
148024 160030: contig of 12007 bp in length  
160031 160130: gap of unknown length  
160131 175441: contig of 15311 bp in length  
175442 175541: gap of unknown length  
175542 191133: contig of 15592 bp in length  
191134 191233: gap of unknown length  
191234 231912: contig of 40679 bp in length.

FEATURES  
source  
1. .231912  
Location/Qualifiers

```

----- Summary Statistics
Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 146479 bases at least Q40
Consensus quality: 149621 bases at least Q30
Consensus quality: 150859 bases at least Q20
Insert size: 151000; agarose-fp
Quality coverage: 4.9 in Q20 bases; agarose-fp
Quality coverage: 4.9 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1986: contig of 1986 bp in length
* 1987 2086: gap of 100 bp
* 2087 5715: contig of 3629 bp in length
* 5716 5815: gap of 100 bp
* 5816 9963: contig of 4148 bp in length
* 9964 10063: gap of 100 bp
* 10064 16895: contig of 6832 bp in length
* 16896 16995: gap of 100 bp
* 16996 32153: contig of 15164 bp in length
* 32160 32259: gap of 100 bp
* 32260 42225: contig of 9966 bp in length
* 42226 42325: gap of 100 bp
* 42326 54505: contig of 12180 bp in length
* 54506 54605: gap of 100 bp
* 54606 71355: contig of 16750 bp in length
* 71356 71455: gap of 100 bp
* 71456 86668: contig of 15213 bp in length
* 86669 86768: gap of 100 bp
* 86769 109678: contig of 22910 bp in length
* 109679 109778: gap of 100 bp
* 109779 152751: contig of 42973 bp in length.
*
* Location/Qualifiers
* 1. 152751
* /organism="Homo sapiens"
* /db_xref="taxon:9606"
* /chromosome="4"
* /map="4"
* /clone="RP11-257M8"
* /clone_lib="RP11-11 Human Male BAC"
*
* 1. 1986
* /note="assembly_fragment"
* 2087. 5715
* /note="assembly_fragment"
* 5816. 9963
* /note="assembly_fragment"
* 10064. 16895
* /note="assembly_fragment"
* 16996. 32153
* /note="assembly_fragment"
* 32260. 42225
* /note="assembly_fragment"
* clone_end:T7
* vector_side:left"
* 42326. 54505
* /note="assembly_fragment"
* 54606. 71355
* /note="assembly_fragment"
* 71456. 86668
* /note="assembly_fragment"
* clone_end:SP6
* vector_side:left"
* 86769. 109678
* /note="assembly_fragment"

```

```

-----
Query Match 40.1% Score 55; DB 2; Length 231912;
Best Local Similarity 66.4%; Pred. No. 0.092;
Matches 79; Conservative 0; Mismatches 40; Indels 0; Gaps 0;
QY 18 GAACATCCACATAAATAAAGCAGAAAAAGAAATGAAAGAAAGATGAGCCAG 77
|||||C| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 12099 GAAAAAAGGAAAAAGGAAAAAGGAAAAAGGAAAAAGGAAAAAGGAGGAAA 12158
QY 78 AGATTGCTATAAGGAGGAGCTGGAGAAAAAGAGACAAAAAGGAAAAAGGAAAAAAG 136
| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 12159 AAAAGGGAAAAAGAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAG 12217

RESULT 14
AC058806/c
LOCUS
DEFINITION Homo sapiens chromosome 4 clone RP11-257M8 map 4, WORKING DRAFT
SEQUENCE, 11 unordered pieces.
AC058806
AC058806.1 GI:7596942
VERSION
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 152751)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 4, clone RP11-257M8
Unpublished
2 (bases 1 to 152751)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavsky,L., Bouckhalter,B., Brown,A., Burkett,G.,
Campotiano,A., Castle,A., Choepel,X., Colangelo,M., Collins,S.,
Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Glinde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrum,J., Meneus,L., Mihoval,T., Miranda,C., Mlenga,V., Morrow,J.,
Murphy,T., Naylor,T.M., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,J., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tessier,S., Theodore,J., Tirrell,A., Travers,M., Trigglio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (19-APR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
-----
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
-----
Project Information
Center project name: L8667
Center clone name: 257_M8

```



Mon Aug 5 11:51:42 2002

us-08-973-363-13.rge

Page 13

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 2, 2002, 20:53:00 ; Search time 1205.34 Seconds  
(without alignments)  
195.146 Million cell updates/sec

Title: US-08-973-363-13

Perfect score: 137

Sequence: 1 GATGGGATTGTTTCAGTGAA.....AAAAGAAAGAAATAAGA 137

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N\_Geneseq\_032802.\*  
1: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	135.4	98.8	137	18	AAT42753
2	121	88.3	265	18	AAT42752
3	52.2	38.1	300	20	AAZ14372
4	52.2	38.1	392	22	AAI88688
5	52	38.0	903	22	AAI87708
6	52	38.0	1708	22	AAF26570
7	52	38.0	2240	21	AAC93391
8	52	38.0	2240	22	AAC08375
9	52	38.0	2240	22	AAS02536

10	51.8	37.8	170	22	AAK84661	Human immune/haema
11	51.8	37.8	2040	22	AAI63879	Human polynucleoti
12	51.8	37.8	2288	22	AAH33398	Human colon cancer
13	51.8	37.8	2288	22	AAF72807	Secreted protein g
14	51.2	37.4	733	22	AAI96873	Human neuroblastom
15	51	37.2	317	22	AAS29066	cDNA encoding for
16	51	37.2	418	22	AAI88883	Human polynucleoti
17	51	37.2	464	22	AAH33241	Human colon cancer
18	51	37.2	1954	21	AAA23441	cDNA encoding huma
19	51	37.2	2510	22	AAH33295	Human colon cancer
20	50.8	37.1	165	22	ABA17044	Human nervous syst
21	50.8	37.1	165	22	ABA08161	Human ovarian and
22	50.8	37.1	165	22	AAS39812	Genomic sequence #
23	50.8	37.1	165	22	AAD16672	Genomic sequence #
24	50.8	37.1	165	22	AAS29235	Human pancreatic r
25	50.8	37.1	165	22	AAI04803	Human reproductive
26	50.8	37.1	165	22	AAI07127	Human immune/haema
27	50.8	37.1	165	22	AAK73944	Human immune/haema
28	50.8	37.1	165	22	AAK78761	Human immune/haema
29	50.8	37.1	165	22	AAK79574	Human immune/haema
30	50.8	37.1	165	22	AAK81312	Human immune/haema
31	50.8	37.1	165	22	AAK83381	Human digestive sy
32	50.8	37.1	165	22	AAK89974	Human digestive sy
33	50.8	37.1	165	22	AAK90168	Human digestive sy
34	50.8	37.1	202	21	AAC98563	Human colon cancer
35	50.8	37.1	597	22	AAH71472	Human cervical can
36	50.6	36.9	1985	21	AAC59406	Human immune syste
37	50.6	36.9	8392	24	ABL33490	Human polynucleoti
38	50.4	36.8	445	22	AAI88613	cDNA encoding huma
39	50.4	36.8	1933	21	AAA23442	Human immune syste
40	50.4	36.8	37973	24	ABL34196	Human immune syste
41	50.2	36.6	1298	22	AAH2764	Human cervical can
42	50.2	36.6	2198	21	AAA23424	cDNA encoding huma
43	50.2	36.6	4167	22	ABA08224	Human ovarian and
44	50.2	36.6	4167	22	AAS29236	Genomic sequence #
45	50.2	36.6	4167	22	AAS30018	Human lung antigen

# ALIGNMENTS

RESULT 1  
AAT42753  
ID AAT42753 standard; cDNA; 137 BP.  
XX  
AC AAT42753;  
XX  
DT 12-MAR-1997 (first entry)  
XX  
DE Chicken CHD-W clone CC14 3' motif.  
XX  
KW Bird; sex determination; chromodomain-Helicase-DNA binding 1 Avian;  
KW CHD-W; chromodomain-Helicase-DNA binding on the W chromosome; ss.  
XX  
OS Gallus sp.  
XX  
FH Key Location/Qualifiers  
FT misc\_difference 52  
FT /tag= a  
FT /note= "base 52 disrupts the reading frame for  
FT the translated amino acid sequence given  
FT in Fig 7"

WO9639505-A1.  
12-DEC-1996.  
05-JUN-1996; 96WO-GB01341.  
06-JUN-1995; 95GB-0011439.  
(ISIS-) ISIS INNOVATION LTD.







```
XX 28-FEB-2000; 2000US-0515126.
PR 18-MAY-2000; 2000US-0577409.
XX (HYSE-) HYSEQ INC.
XX Tang YT, Liu C, Drmanac RT;
PI WPI; 2001-514838/56.
XX P-PSDB; AA007777.
DR Isolated nucleic acids and polypeptides, useful for preventing
XX diagnosing and treating e.g. leukaemia, inflammation and immune
PT disorders -
PT Claim 1; SEQ ID NO 7768; 1399pp + Sequence Listing; English.
XX CC The invention relates to human polynucleotides (AA179941-AA193841) and
XX the encoded proteins (AA000010-AA013910) that exhibit activity elating to
XX cytokine, cell proliferation or cell differentiation or which may induce
XX production of other cytokines in other cell populations. The
XX polynucleotides and polypeptides are useful in gene therapy, vaccines or
XX peptide therapy. The polypeptides have various cytokine-like activities,
XX e.g. stem cell growth factor activity, haematopoiesis regulating
XX activity, tissue growth factor activity, immunomodulatory activity and
XX activin/inhibin activity and may be useful in the diagnosis and/or
XX treatment of cancer, leukaemia, nervous system disorders, arthritis and
XX inflammation.
XX CC Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 903 BP; 297 A; 208 C; 192 G; 199 T; 7 other;

Query Match 38.0%; Score 52; DB 22; Length 903;
Best Local Similarity 62.1%; Pred. No. 0.041;
Matches 82; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 6 GATTGTTTCAGTGAAACATCCACATAAAATAAAGCAGAAAAAAGAAATGAAGAAA 65
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 211 gaatttttaagtaaaaaaataaaaaaataaaaaaataaaaaaataaaaaaata 270
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 66 AAGATGAGCCAGAGATTGTGTATTAAGAGAGAGCTGGAGAAAAAGAGACAAAAAGAAA 125
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 271 aaaaaaataaaaaaataaaaaaataaaaaaataaaaaaataaaaaaataaaaaa 330
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 126 AGCAAAATAGA 137
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 331 aaaaaaataaaaaa 342

RESULT 6
AAF26570
ID AAF26570 standard; DNA; 1708 BP.
XX AC AAF26570;
XX 27-MAR-2001 (first entry)
DT DNA encoding human secreted protein #24.
DE Secreted protein; gene therapy; vaccine; cancer; leukemia;
XX autoimmune disease; allergy; inflammation; graft rejection;
KW hyperproliferation; cardiovascular; infection; ss.
XX OS Homo sapiens.
XX PN WO200076531-A1.
XX PD 21-DEC-2000.
XX PF 01-JUN-2000; 2000WO-US15137.
PR
```

```
XX 11-JUN-1999; 99US-0138625.
XX (HUMA-) HUMAN GENOME SCI INC.
XX PA Rosen CA, Ruben SM, Komatsoulis GA;
XX WPI; 2001-071148/08.
XX DR Nucleic acids encoding 47 human secreted polypeptides, useful for
XX preventing, diagnosing and/or treating e.g. cancers, Parkinson's
PT disease and diabetic retinopathy -
PT Claim 1; Page 458-459; 525pp; English.
XX CC The present invention relates to 26 secreted human proteins. The
XX proteins may be used in the prevention, diagnosis and treatment of
XX diseases associated with inappropriate polypeptide expression.
XX For example, they may be used in gene therapy or in vaccines.
XX Typical of diseases which are potentially treatable are cancers
XX (including leukemia), autoimmune diseases, allergies, inflammation,
XX graft rejection, hyperproliferation, cardiovascular diseases
XX (particularly critical limb ischemia and coronary disease) and any
XX involving abnormal angiogenesis, neurodegeneration and/or
XX infectious diseases.
XX SQ Sequence 1708 BP; 410 A; 425 C; 474 G; 398 T; 1 other;

Query Match 38.0%; Score 52; DB 22; Length 1708;
Best Local Similarity 62.1%; Pred. No. 0.041;
Matches 82; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 6 GATTGTTTCAGTGAAACATCCACATAAAATAAAGCAGAAAAAAGAAATGAAGAAA 65
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1576 gtggctgtgtgaaaaaataaaaaaataaaaaaataaaaaaataaaaaaataaaa 1635
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 66 AAGATGAGCCAGAGATTGTGTATTAAGAGAGAGCTGGAGAAAAAGAGACAAAAAGAAA 125
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1636 aaaaaaataaaaaaataaaaaaataaaaaaataaaaaaataaaaaaataaaaaa 1695
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 126 AGCAAAATAGA 137
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1696 aaaaaaataaaaaa 1707

RESULT 7
AAC93391
ID AAC93391 standard; cDNA; 2240 BP.
XX AC AAC93391;
XX 26-FEB-2001 (first entry)
DT Human secreted protein gene 28 SEQ ID NO:38.
DE Human; secreted protein; diagnosis; immunosuppressive; antiarthritic;
XX antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic;
KW cerebroprotective; nootropic; neuroprotective; antibacterial; virucide;
KW fungicide; ophthalmological; gene therapy; autoimmune disease; infection;
KW hyperproliferative disorder; cardiovascular disorder; angiogenesis;
KW cerebrovascular disorder; nervous system disorder; ocular disorder;
KW wound healing; skin aging; food additive; preservative; ss.
XX OS Homo sapiens.
XX PN WO200061620-A1.
XX PD 19-OCT-2000.
XX PF 06-APR-2000; 2000WO-US09069.
XX PR 09-APR-1999; 99US-0128702.
PR
```

PR 20-JAN-2000; 2000US-0177049.  
XX (HUMA-) HUMAN GENOME SCI INC.  
PA (ROSE/) ROSEN C A.  
XX Rosen CA, Ruben SM, Komatsoulis G;  
XX WPI; 2000-619225/59.  
DR P-PSDB; AAB51647.  
XX  
XX Isolated nucleic acid molecule encoding a human secreted protein is  
PT used in preventing, treating or ameliorating a medical condition -  
PT  
PS Claim 1; Page 443-444; 540pp; English.  
XX  
XX The polynucleotide sequences given in AAC93364 to AAC93412 encode the  
CC human secreted proteins given in AAB51620 to AAB51668. AAB51669 to  
CC AAB51722 represent human secreted polypeptide sequences and proteins  
CC homologous to them, which are given in the exemplification of the present  
CC invention. Human secreted proteins have activities based on the tissues  
CC and cells the genes are expressed in. Examples of activities include:  
CC antiarthritic; immunosuppressive; antirheumatic; antiproliferative;  
CC cytoskeletal; cardiant; vasotropic; cerebroprotective; neurotropic;  
CC neuroprotective; antibacterial; virucide; fungicide; and  
CC opthalmological. The polynucleotides and polypeptides can be used to  
CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,  
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used  
CC in diagnosing a pathological condition or susceptibility to a  
CC pathological condition. Disorders which are diagnosed or treated include  
CC autoimmune diseases, hyperproliferative disorders, cardiovascular  
CC disorders, cerebrovascular disorders, angiogenesis, nervous system  
CC disorders, infections caused by bacteria, viruses and fungi and ocular  
CC disorders. The polypeptides can also be used to aid wound healing and  
CC epithelial cell proliferation, to prevent skin aging due to sunburn, to  
CC maintain organs before transplantation, for supporting cell culture of  
CC primary tissues, to regenerate tissues and in chemotaxis. The  
CC polypeptides can also be used as a food additive or preservative to  
CC increase or decrease storage capabilities. AAC93355 to AAC93363 and  
CC AAB51619 represent sequences which are used in the exemplification of the  
XX present invention.  
SQ Sequence 2240 BP; 532 A; 552 C; 623 G; 531 T; 2 other;

Query Match 38.0%; Score 52; DB 21; Length 2240;  
Best Local Similarity 62.1%; Pred. No. 0.041; 50; Indels 0; Gaps 0;  
Matches 82; Conservative 0; Mismatches 0;  
QY 6 GATTGTTTCAGTGAACATCCATATAAATAAATGAAGCAGAAAAAAGAAATGAAGAAA 65  
Db 2108 gtctgtctgtgagtgaaaaaataaaaaaataaaaaaataaaaaaataaaaaa 2167  
QY 66 AGATGAGCGCAGATGTTGATATAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 125  
Db 2168 aaaaaaataaaaaaataaaaaaataaaaaaataaaaaaataaaaaaataaaaaa 2227  
QY 126 AGGAAATAAGA 137  
Db 2228 aaaaaaagaaaa 2239  
RESULT 8  
AAD08375  
ID AAD08375 standard; cDNA; 2240 BP.  
AC  
XX AAD08375;  
XX  
DT 09-AUG-2001 (first entry)  
XX  
DE Human secreted protein-encoding gene 31 cDNA clone HTLHN86, SEQ ID NO:41.  
XX Human; secreted protein; proliferative disorder; cancer; tumour;  
KW foetal abnormality; developmental abnormality; haematopoietic disorder;  
XX

KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;  
KW inflammation; allergy; neurological disorder; Alzheimer's disease;  
KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;  
KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;  
KW cardiovascular disorder; angioecnic disorder; kidney disorder;  
KW gastrointestinal disorder; pregnancy-related disorder;  
KW endocrine disorder; infection; wound healing; vulnery;  
KW cell culture; chemotaxis; food additive; gene therapy;  
KW binding partner identification; ss.  
XX  
XX Homo sapiens.  
XX  
XX Key Location/Qualifiers  
FH 1413..1976  
FT /tag= a  
FT /product= "human secreted protein precursor"  
FT 1413..1472  
FT /tag= b  
FT 1473..1973  
FT mat\_peptide  
FT /tag= c  
FT /product= "Mature human secreted protein"  
XX WO2000077022-A1.  
PN  
XX  
XX 21-DEC-2000.  
XX  
XX 01-JUN-2000; 2000WO-US15136.  
XX  
XX 11-JUN-1999; 99US-0138629.  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Ruben SM, Komatsoulis GA;  
PI P-PSDB; AAE03928.  
DR WPI; 2001-367020/38.  
XX  
XX Nucleic acids encoding 50 human secreted polypeptides, useful for  
PT preventing, diagnosing and/or treating diseases, e.g. Parkinson's  
PT disease, botulism, cancers and Scimitar syndrome -  
XX  
XX Claim 1; Page 511-512; 614pp; English.  
XX  
XX AAD08345-AAD08394 represent cDNAs corresponding to 50 human secreted  
CC protein genes and AAE03898-AAE03947 represent the proteins they encode.  
CC AAE03948-AAE03996 represent human secreted protein fragments or variants.  
CC The genes and their secreted proteins are useful for preventing,  
CC treating or ameliorating medical conditions, e.g., by protein or gene  
CC therapy. Pathological conditions can be diagnosed by determining the  
CC amount of the new protein in a sample or by determining the presence of  
CC mutations in the new genes. Specific uses are described for each of the  
CC 50 genes, based on the tissues in which they are most highly expressed,  
CC and include developing products for the diagnosis or treatment of  
CC proliferative disorders, cancer, tumours, foetal and developmental  
CC abnormalities, haematopoietic disorders, diseases of the immune system,  
CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,  
CC allergies, neurological disorders (e.g., Alzheimer's disease,  
CC Parkinson's disease), cognitive disorders, schizophrenia, asthma,  
CC skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,  
CC cardiovascular disorders, angioecnic disorders, kidney disorders,  
CC gastrointestinal disorders, pregnancy-related disorders, endocrine  
CC disorders, and infections. The proteins can also be used to aid wound  
CC healing and epithelial cell proliferation to prevent skin aging due to  
CC sunburn, to maintain organs before transplantation, for supporting cell  
CC culture of primary tissues, to regenerate tissues, to identify their  
CC cognate ligands or binding partners, and in chemotaxis, and can be used  
CC as a food additive or preservative to modify storage properties.  
CC Antibodies specific for a protein of the invention can be used in  
CC alleviating symptoms associated with the disorders mentioned above, and  
CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked  
CC immunosorbent assay (ELISA). The present sequence represents a human  
CC secreted protein-encoding cDNA of the invention.  
XX

```
SQ Sequence 2240 BP; 532 A; 552 C; 622 G; 531 T; 3 other;

Query Match      38.0%; Score 52; DB 22; Length 2240;
Best Local Similarity 62.1%; Pred. No. 0.041;
Matches 82; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 6 GATTGTTTCAGTGAACATCCACATATAAAATAAAGCAGAAAAAATGAAGAAA 65
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2108 gttggtcgtgagtgagagagagagagagagagagagagagagagagagag 2167
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 66 AAGATGAGCCAGAGATTGGTATATAAGAGAGCTGGAGAAAAAGAGACACAAAAAGAAA 125
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2168 aaaaaaaagagagagagagagagagagagagagagagagagagagagagagag 2227
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 126 AGGAAAAATAAGA 137
   | | | | | | | |
Db 2228 aaaaaaagagaaa 2239

RESULT 9
AAS02536
ID AAS02536 standard; cDNA; 2240 BP.
XX
AC AAS02536;
DT 18-JUL-2001 (first entry)
DE Human secreted protein gene #17.
XX
KW Human secreted protein; diagnosis; autoimmune disease; ss;
KW rheumatoid arthritis; hyperproliferative disorder; neoplasm; sunburn;
KW cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
KW cerebral ischaemia; angiogenesis; nervous system disorder; skin aging;
KW Alzheimer's disease; infection; ocular disorder; corneal infection;
KW wound healing; epithelial cell proliferation; chemotaxis; preservative;
KW organ transplantation; tissue regeneration; food additive.
OS Homo sapiens.
XX
PN WO200123409-A2.
XX
PD 05-APR-2001.
XX
PF 26-SEP-2000; 2000WO-US26371.
XX
PR 27-SEP-1999; 99US-0155804.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM, Komatsoulis GA;
XX
DR WPI; 2001-266139/27.
DR P-PSDB; AAU01675.
XX
PT Nucleic acids encoding 38 human secreted polypeptides, useful for
PT preventing, diagnosing and/or treating e.g. cancers, Parkinson's
PT disease and diabetic retinopathy -
XX
PS Disclosure; Page 429; 488pp; English.
XX
XX
AAS02511-AAS02557 represent human secreted protein coding sequences
CC and primers of the invention. The human secreted protein sequences are
CC used to prevent, treat or ameliorate a medical condition in e.g. humans,
CC mice, rabbits, goats, horses, cats, dogs, chickens or sheep. They are
CC also used in diagnosing a pathological condition or susceptibility to a
CC pathological condition. The antibodies to human secreted proteins can
CC also be used in alleviating symptoms associated with the disorders and in
CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked
CC immunosorbent assays (ELISA). Disorders which are diagnosed or treated
CC include autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative
CC disorders e.g. neoplasms of the breast or liver, cardiovascular disorders
CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
CC
```

```
CC anglogenesis, nervous system disorders e.g. Alzheimer's disease,
CC infections caused by bacteria, viruses and fungi and ocular disorders
CC e.g. corneal infection. The polypeptides can also be used to aid wound
CC healing and epithelial cell proliferation, to prevent skin aging due to
CC sunburn, to maintain organs before transplantation, for supporting cell
CC culture of primary tissues, to regenerate tissues and in chemotaxis. The
CC polypeptides can also be used as a food additive or preservative to
CC increase or decrease storage capabilities.
XX
SQ Sequence 2240 BP; 532 A; 552 C; 623 G; 531 T; 2 other;

Query Match      38.0%; Score 52; DB 22; Length 2240;
Best Local Similarity 62.1%; Pred. No. 0.041;
Matches 82; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 6 GATTGTTTCAGTGAACATCCACATATAAAATAAAGCAGAAAAAATGAAGAAA 65
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2108 gttggtcgtgagtgagagagagagagagagagagagagagagagagagag 2167
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 66 AAGATGAGCCAGAGATTGGTATATAAGAGAGCTGGAGAAAAAGAGACACAAAAAGAAA 125
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2168 aaaaaaaagagagagagagagagagagagagagagagagagagagagagagag 2227
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 126 AGGAAAAATAAGA 137
   | | | | | | | |
Db 2228 aaaaaaagagaaa 2239

RESULT 10
AAS02536
ID AAS02536 standard; DNA; 170 BP.
XX
AC AAS02536;
XX
DT 07-NOV-2001 (first entry)
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:39473.
XX
DE Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS Homo sapiens.
XX
PN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01354.
XX
XX
31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
```

PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0227182.  
PR 22-AUG-2000; 2000US-0227829.  
PR 23-AUG-2000; 2000US-0227829.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229387.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234597.  
PR 25-SEP-2000; 2000US-0234988.  
PR 26-SEP-2000; 2000US-0235488.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX PI Rosen CA, Barash SC, Ruben SM;  
XX WPI: 2001-483426/52.  
XX  
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
XX useful for preventing, diagnosing and/or treating cancers and  
XX metastasis.  
XX  
XX Disclosure: SEQ ID NO 39473; 3071pp + Sequence Listing: English.  
XX  
XX AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)  
XX amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic  
XX activity, and can be used in gene therapy and vaccine production. (I)  
XX proteins and polynucleotides may be used in the prevention, diagnosis and  
XX treatment of diseases associated with inappropriate (I) expression. For  
XX example, they may be used to treat disorders associated with decreased  
XX expression by rectifying mutations or deletions in a patient's genome  
XX that affect the activity of (I) by expressing inactive proteins or to  
XX supplement the patients' own production of (I). Additionally, (I)  
XX polynucleotides may be used to produce the secreted (I), by inserting the  
XX nucleic acids into a host cell and culturing the cell to express the  
XX protein. (I) proteins and polynucleotides may be used to prevent,  
XX diagnose and treat immune/hematopoietic-related diseases, especially  
XX cancers and cancer metastases of hematopoietic-derived cells. AAK64703  
XX to AAK87694 represent human immune/hematopoietic antigen genomic  
XX sequences from the present invention. AAK54942 to AAK54950 and AAK82169  
XX represent sequences used in the exemplification of the present invention.  
XX  
XX Sequence 170 BP; 165 A; 0 C; 4 G; 1 T; 0 other;

```
Query Match          37.8%; Score 51.8; DB 22; Length 170;
Best Local Similarity 64.7%; Pred. No. 0.044;
Matches 77; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

QY 19 AACATCCACATAAAATAAGCAGAAAAAGAAATGAAGAAAAAGATGAGCCAGA 78
    ||| | | ||||| |||| | |||| | |||| | |||| | |||| | |||| |
Db 12 aaaaaaaataaaaaaaagaaaaaaagaaaaaaagaaaaaaagaaaaaaagaaa 71
    | | | | | ||| | | | | | | | | | | | | | | | | | | | | |
QY 79 GATTGGTATTAAGAGAACTGGAGAAAAAGAGACAAAAAGAAAAATAAGA 137
    | | | | | ||| | | | | | | | | | | | | | | | | | | | | |
Db 72 aaaaaaaataaaaaaaagaaaaaaagaaaaaaagaaaaaaagaaaaaaagaaa 130

RESULT 11
AAI63879
ID AAI63879 standard; cDNA; 2040 BP.
XX
AC AAI63879;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 87.
XX
KW Human; antiarthritic; antirheumatic; antiproliferative; vasotropic;
KW cerebroprotective; nootropic; neuroprotective; antibacterial; virucide;
KW fungicide; ophthalmological; cytostatic; immunosuppressive; nootropic;
KW neuroprotective; antiallergic; hepatotropic; antidiabetic;
KW antiinflammatory; antiulcer; vulnery; anticonvulsant; antibacterial;
KW antiparasitic; cardiant; gene therapy; cancer; immune disorder;
KW cardiovascular disorder; neurological disease; infection; human; ss.
XX
OS Homo sapiens.
XX
PN WO200155308-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01309.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR- 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
```

Qy	71	GAGCCAGAGATTGCTATTAAGAAGACGTGGAGAAAAAAGAACAGACAGACAAGAAAGCGRA	130
Db	1956	aaaaaaatttatcttaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa	2015
Qy	131	ANTAGA 137	
Db	2016	aaaaaa 2022	
RESULT	12		
AAH33398			
ID	AAH33398	standard; cDNA; 2288 BP.	
XX	AAH33398;		
AC			
XX			
DT	03-SEP-2001	(first entry)	
XX			
DE	Human colon cancer antigen encoding cDNA SEQ ID NO:454.		
XX			
KW	Human; colon cancer; colon cancer antigen; diagnosis; detection;		
KW	colorectal carcinoma; ss.		
OS	Homo sapiens.		
PN	WO200122920-A2.		
PD	05-APR-2001.		
PF	28-SEP-2000; 2000WO-US26524.		
PR	29-SEP-1999; 99US-0157137.		
PR	03-NOV-1999; 99US-0163280.		
PA	(HUMA-) HUMAN GENOME SCI INC.		
PI	Ruben SM, Barash SC, Birse CE, Rosen CA;		
XX			
WI	WPI: 2001-235357/24.		
DR	P-FSDB; AAG73967.		
PT	Nucleic acids encoding 4277 human colon cancer-associated polypeptides,		
XX	useful for preventing, diagnosing and/or treating colorectal cancers -		
PS	Claim 1; Page 2549; 9803pp; English.		
XX			
CC	AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon		
CC	cancer-associated nucleic acid molecules (N) and proteins (P), where		
CC	the proteins are collectively known as colon cancer antigens, where		
CC	cancer antigens have cytostatic activity and can be used in gene		
CC	therapy and vaccine production. N and P may be used in the prevention,		
CC	diagnosis and treatment of diseases associated with inappropriate p		
CC	expression. For example, N and P may be used to treat disorders		
CC	associated with decreased expression by rectifying mutations or deletions		
CC	in a patient's genome that affect the activity of p by expressing		
CC	inactive proteins or to supplement the patients own production of P.		
CC	Additionally, N may be used to produce the colon cancer-associated ps,		
CC	by inserting the nucleic acids into a host cell and culturing the cell		
CC	to express the proteins. N and P can be used in the prevention, diagnosis		
CC	and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204		
CC	and AAB77789 represent sequences used in the exemplification of the		
CC	present invention.		
CC	N.B. Pages 666 to 682 and page 7053 of the sequence listing were		
CC	missing at time of publication, meaning no sequences are present for		
CC	SEQ ID NO:1027 to 1052, 7921 and 7922.		
XX			
SQ	Sequence 2288 BP; 718 A; 467 C; 545 G; 558 T; 0 other;		
XX			

Query Match	37.8%;	Score 51.8;	DB 22;	Length 2288;
Best Local Similarity	63.0%;	Pred. No. 0.045;		
Matches 80;	Conservative	0;	Mismatches 47;	Indels 0;
QV	11	TTTTTCAGTGAAGACATCCACAT	AAAAAAAAATTAAGACGAGAAAAAGAAATTAAGAAAAAGAT	70

```

Db 2154 ttctactgaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2213
QY 71 GAGCCAGAGATTGGTATAAGAGAGGAGCTGGAGAGAAAAGAGAGACAAAGAAAAGGAA 130
Db 2214 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2273
QY 131 AATAAGA 137
Db 2274 aaaaaaa 2280

RESULT 13
AAF72807
ID AAF72807 standard; DNA; 2288 BP.
XX
AC AAF72807;
XX
DT 24-APR-2001 (first entry)
XX
DE Secreted protein gene #9.
XX
KW Secreted protein; human; autoimmune; hyperproliferation;
KW cardiovascular; cerebrovascular; infection; food; ds.
XX
OS Homo sapiens.
XX
PN WO200107459-A1.
XX
XX
PD 01-FEB-2001.
XX
PF 20-JUL-2000; 2000WO-US19735.
XX
PR 23-JUL-1999; 99US-0145220.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM, Ebner R, Duan RD, Ni J, Soppet DR, Moore PA;
PI Shi Y, Lafleur DW, Olsen HS, Birse CE, Komatsoulis GA;
XX
XX WPI; 2001-123261/13.
XX
XX New isolated nucleic acid encoding 29 secreted proteins, for
PT diagnosing, preventing and treating e.g. autoimmune,
PT hyperproliferative, cardiovascular, and ocular diseases or disorders
PT and microorganism infections -
XX
PS Claim 1; Page 499-500; 60lpp; English.
XX
CC The present invention relates to 29 human secreted proteins. The
CC invention is used to prevent autoimmune diseases e.g. rheumatoid
CC arthritis, hyperproliferative disorders e.g. neoplasms of the
CC breast or liver, cardiovascular disorders e.g. cardiac arrest,
CC cerebrovascular disorders e.g. cerebral ischemia, angiodenesis,
CC nervous system disorders e.g. Alzheimer's disease, infections
CC caused by bacteria, viruses and fungi and ocular disorders e.g.
CC corneal infection. Also used in food preparations.
XX
SQ Sequence 2288 BP; 718 A; 467 C; 545 G; 558 T; 0 other;

Query Match 37.8%; Score 51.8; DB 22; Length 2288;
Best Local Similarity 63.0%; Pred. No. 0.045;
Matches 80; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 11 TTTCAGTGAACATCCACATAAAAAAATAAAGCAGAAAAAGAAAATGAAGAAAAGAT 70
Db 2154 ttctactgaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2213
QY 71 GAGCCAGAGATTGGTATAAGAGAGGAGCTGGAGAGAAAAGAGAGACAAAGAAAAGGAA 130
Db 2214 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2273
```

```

QY 131 AATAAGA 137
Db 2274 aaaaaaa 2280

RESULT 14
AAI96873/C
ID AAI96873 standard; cDNA; 733 BP.
XX
AC AAI96873;
XX
DT 13-NOV-2001 (first entry)
XX
DE Human neuroblastoma expressed polynucleotide SEQ ID NO 2948.
XX
KW Human; neuroblastoma; malignancy; cancer; tumour marker; N-myc; TrkA; ss.
XX
OS Homo sapiens.
XX
PN WO200166719-A1.
XX
PD 13-SEP-2001.
XX
PF 02-MAR-2001; 2001WO-JP01629.
XX
PR 07-MAR-2000; 2000JP-0159195.
XX
PA (CHIB-) CHIBA PREFECTURE.
PA (HISM) HISAMITSU PHARM CO LTD.
XX
PI Nakagawara A;
XX
XX WPI; 2001-565584/63.
XX
PT Nucleic acids originating in gene expressed in human neuroblastoma,
PT useful as probe or primer in diagnosing prognosis of human
PT neuroblastoma, malignancy and susceptibility indicator or tumour marker
PT for anti-cancer agents -
XX
PS Claim 1; Page 2152; 2979pp; Japanese.
XX
CC The invention relates to novel genes (AAI93926-AAI97963) expressed in
CC human neuroblastoma. The nucleic acids are applicable as a probe or
CC primer in diagnosing the prognosis of human neuroblastoma, malignancy and
CC susceptibility indicators or tumour markers for anti-cancer agents. The
CC gene information for diagnosing prognosis is related to factors similar
CC to that for N-myc and TrkA genes.
XX
SQ Sequence 733 BP; 127 A; 155 C; 170 G; 238 T; 43 other;

Query Match 37.4%; Score 51.2; DB 22; Length 733;
Best Local Similarity 66.4%; Pred. No. 0.059;
Matches 71; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 26 CACATAAAAAATAAAGCAGAAAAAAGAAAATGAAGAAAAGATGAGCCAGAGATTGGT 85
Db 131 CNCNAAAAAAGAAAAAAGAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA 72
QY 86 ATAAGAAGGAAGCTGGAGAAAAAAGAGACAAAAAGAGAAAAAGGAAAA 132
Db 71 AAAAAAAGAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA 25

RESULT 15
AAS29066
ID AAS29066 standard; cDNA; 317 BP.
XX
AC AAS29066;
XX
DT 21-NOV-2001 (first entry)
XX
DE cDNA encoding for human DNA-binding protein #37.
```



XX Human; DNA-binding protein; histone; chromo domain protein;  
KW chromatin organisation modifier; Y-box binding protein;  
KW DNA organisation; gene transcription; malignant disease;  
KW autoimmune disorder; rheumatic disease; genetic abnormality;  
KW infectious disease; neurological disorder; gene therapy;  
KW immunomodulatory; anti-HIV; anti rheumatic; anti microbial;  
KW cytostatic; ss.  
XX  
OS Homo sapiens.  
XX  
PW WO200155162-A1.  
XX  
XX  
PD 02-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US01305.  
XX  
XX 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.

PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234957.  
PR 25-SEP-2000; 2000US-0234958.  
PR 26-SEP-2000; 2000US-0234984.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 13-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249284.  
PR 17-NOV-2000; 2000US-0249285.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 2, 2002, 16:53:17 ; Search time 8940.62 Seconds  
(without alignments)  
206.818 Million cell updates/sec

Title: US-08-973-363-13

Sequence: 1 GATGGGATGTTTCAGTGAA.....AAAGAGAAAGGAATAAAGA 137

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST: \*  
1: em\_estba: \*  
2: em\_esthum: \*  
3: em\_estin: \*  
4: em\_estmu: \*  
5: em\_estov: \*  
6: em\_estpl: \*  
7: em\_estro: \*  
8: em\_hic: \*  
9: gb\_est1: \*  
10: gb\_est2: \*  
11: gb\_hic: \*  
12: gb\_hic: \*  
13: em\_gss\_hum: \*  
14: em\_gss\_inv: \*  
15: em\_gss\_pin: \*  
16: em\_gss\_vit: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	90.6	66.1	821	10	BF239967 601905170
2	83.4	60.9	343	9	AA748563 0a57c09.s
3	83.4	60.9	430	9	AI242163 0h81f02.s
4	83.4	60.9	456	9	AA699918 2i61f12.s
5	83.4	60.9	677	10	BM015597 603641828
6	81	59.1	547	9	AI890775 0h90775
7	73.8	53.9	249	9	AI890775 0h90775
8	58.2	42.5	782	10	BF906261 601502247
9	57.8	42.2	973	10	BM415214 0p20286 M
10	56.2	41.0	200	10	BF548029 602575943
11	56	40.9	230	9	AI013222 0c77e07.y
12	55.6	40.6	240	10	BM480932 532632 MA
13	55.2	40.3	656	12	AG040706 Pan trogl
14	55.2	40.3	940	10	BM415207 0p20279 M
15	55	40.1	338	12	CNS006AG
16	55	40.1	457	12	AG040672 Pan trogl
17	55	40.1	467	9	AV685046

18	55	40.1	815	12	AG044049
19	54.6	39.9	449	12	CNS020YQ
20	54.4	39.7	391	9	AU162571
21	54.4	39.7	774	12	AG056519
22	54.2	39.6	214	9	AW100661
23	54.2	39.6	311	9	AL513907
24	54.2	39.6	366	9	AL514543
25	54.2	39.6	508	9	AV681993
26	54.2	39.6	533	12	AZ507189
27	54.2	39.6	728	9	AL569904
28	54.2	39.6	865	12	CNS04H64
29	54.2	39.6	1083	10	BE965014
30	54	39.6	1083	12	BL1974
31	54	39.4	938	10	BM415992
32	53.8	39.3	161	10	BM128138
33	53.8	39.3	301	12	CNS03C4J
34	53.8	39.3	313	12	CNS00CZT
35	53.8	39.3	710	12	CNS01WUM
36	53.8	39.3	717	12	CNS0181Z
37	53.8	39.3	849	12	AG044091
38	53.8	39.3	1038	12	CNS0166P
39	53.8	39.3	1058	12	CNS0166P
40	53.6	39.1	209	9	AA638753
41	53.6	39.1	857	9	AL550892
42	53.4	39.0	348	10	BG496803
43	53.4	39.0	354	9	AL662542
44	53.4	39.0	406	12	AQ431852
45	53.4	39.0	501	10	BG623840

## ALIGNMENTS

RESULT 1  
LOCUS BF239967 821 bp mRNA linear EST 14-NOV-2000  
DEFINITION 601905170F1 NIH\_MGC\_54 Homo sapiens cDNA clone IMAGE:4133129 5', mRNA sequence.  
ACCESSION BF239967  
VERSION BF239967.1 GI:11153890  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE 1 (bases 1 to 821)  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
TITLE NIH-MGC http://mgc.nci.nih.gov/  
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cga@pds-remail.nih.gov  
Tissue Procurement: ATCC  
CDNA Library Preparation: CLONETECH Laboratories, Inc.  
CDNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)  
CDNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov  
plate: LCM1033 row: k column: 18  
High quality sequence stop: 562.  
Location/Qualifiers  
1. 821  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4133129"  
/tissue="type="from chronic myelogenous leukemia"  
/lab\_host="DH10B (TI phage-resistant)"  
/note="Organ: bone marrow; Vector: pDNR-LIB (Clontech); Site:1: Sfii (ggccattatggcc); Site:2: Sfii (ggccattatggcc); Site:3: Sfii (ggccattatggcc); Double-stranded cDNA was prepared from cell line RNA. 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGCCATTATGCC-3' and 3' adaptor



BASE COUNT 111 a 89 c 50 g 180 t  
 ORIGIN constructed by Bento Soares and M.Fatima Bonaldo."

Query Match 60.9%; Score 83.4; DB 9; Length 430;  
 Best Local Similarity 80.3%; Pred. No. 0.049; 26; Indels 1; Gaps 1;  
 Matches 110; Conservative 0; Mismatches 26; Indels 1; Gaps 1;  
 QY 1 GATGGGATGTTTCAGTGAACATCCATATAAATAAAGCAGAGAAAAGAAATGA 60  
 Db 206 GATGAGATGATCTGTGTAACATCCAAATAAATAAAGCAGAGAG-AGACAGTGA 148  
 QY 61 AGAAAAGATGAGCCAGAGATGTTGTTATTAAGAGAGAGTGGAGAAAAGAGAGACAAA 120  
 Db 147 AGAAAACCTGAGCCAGATGTTTATATAAGAGAGAACAGAGAAAAGAGAGGAGCAA 88  
 QY 121 AGAAAAGAAAATAAGA 137  
 Db 87 AGAAAAGGAGATAAAA 71

RESULT 4  
 AA699918/c  
 LOCUS 456 bp mRNA linear EST 19-DEC-1997  
 DEFINITION z161f12.s1 Soares fetal\_liver spleen\_INFLS\_s1 Homo sapiens CDNA  
 Clone IMAGE:435311 3', mRNA sequence.

ACCESSION AA699918  
 VERSION 1 GI:2702881

KEYWORDS EST.  
 SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.

AUTHORS Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,  
 Krieman, D., Kucaba, I., Lacy, M., Le, N., Lennon, G., Marra, M., Martin,  
 J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B.,  
 White, T., Wylie, T., Waterston, R. and Wilson, R.

WashU-NCI human EST Project  
 Unpublished (1997)  
 Contact: Wilson RK  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@wustl.edu

This clone is available royalty-free through LLNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 Seq primer: -40ml3 fwd. ET from Amersham  
 High quality sequence stop: 420.

FEATURES  
 source  
 1. 456  
 /organism="Homo sapiens"  
 /db\_xref="GDB:1335080"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:435311"  
 /clone\_lib="Soares\_fetal\_liver\_spleen\_INFLS\_s1"  
 /sex="male"  
 /dev\_stage="20 week-post conception fetus"  
 /lab\_host="DH10B (ampicillin resistant)"  
 /note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)  
 with a modified polylinker; Site\_1: Pac I; Site\_2: Eco RI;  
 This is a subtracted version of the original Soares fetal  
 liver spleen INFLS library. 1st strand cDNA was primed  
 with a Pac I - oligo(dn) primer [5'  
 AACTGGAAGAATTAATAAGATCTTTTTTTTTTTTTTTTTTTT 3'],  
 double-stranded cDNA was ligated to Eco RI adaptors  
 (Pharmacia), digested with Pac I and cloned into the Pac I  
 and Eco RI sites of the modified pT7T3 vector. Library  
 went through one round of normalization. Library  
 constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 119 a 95 c 57 g 185 t

ORIGIN

Query Match 60.9%; Score 83.4; DB 9; Length 456;  
 Best Local Similarity 80.3%; Pred. No. 0.047;  
 Matches 110; Conservative 0; Mismatches 26; Indels 1; Gaps 1;  
 QY 1 GATGGGATGTTTCAGTGAACATCCATATAAATAAAGCAGAGAAAAGAAATGA 60  
 Db 202 GATGAGATGATCTGTGTAACATCCAAATAAATAAAGCAGAGAG-AGACAGTGA 144  
 QY 61 AGAAAAGATGAGCCAGAGATGTTGTTATTAAGAGAGAGTGGAGAAAAGAGAGACAAA 120  
 Db 143 AGAAAACCTGAGCCAGATGTTTATATAAGAGAGAACAGAGAAAAGAGAGGAGCAA 84  
 QY 121 AGAAAAGAAAATAAGA 137  
 Db 83 AGAAAAGGAGATAAAA 67

RESULT 5

BM015597  
 LOCUS 677 bp mRNA linear EST 30-OCT-2001  
 DEFINITION 603641828F1 NIH\_MGC\_87 Homo sapiens cDNA clone IMAGE:5417887 5',  
 mRNA sequence.

ACCESSION BM015597  
 VERSION 1 GI:16529951

KEYWORDS EST.  
 SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.

AUTHORS NIH-MGC http://mgs.nci.nih.gov/.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs@mail.nih.gov

Tissue Procurement: DCTD/DP  
 cDNA Library Preparation: Life Technologies, Inc.  
 DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: LLAM12066 row: 9 column: 08  
 High quality sequence stop: 674.

FEATURES  
 source  
 1. 677  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5417887"  
 /clone\_lib="NIH\_MGC\_87"  
 /tissue\_type="mammary adenocarcinoma, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: breast; Vector: pCMV-SPORT6; Site\_1: NotI;  
 Site\_2: SalI; Cloned unidirectionally; oligo-dT primed.  
 Average insert size 1.383 kb. Library enriched for  
 full-length clones and constructed by Life Technologies.  
 Note: this is a NIH\_MGC Library."

BASE COUNT 305 a 92 c 153 g 127 t

ORIGIN

Query Match 60.9%; Score 83.4; DB 10; Length 677;  
 Best Local Similarity 80.3%; Pred. No. 0.04;  
 Matches 110; Conservative 0; Mismatches 26; Indels 1; Gaps 1;  
 QY 1 GATGGGATGTTTCAGTGAACATCCACATAAATAAAGCAGAGAAAAGAAATGA 60  
 Db 90 GATGAGATGATCTGTGTAACATCCAAATAAATAAAGCAGAGAG-AGACAGTGA 148  
 QY 61 AGAAAAGATGAGCCAGAGATGTTGTTATTAAGAGAGAGTGGAGAAAAGAGAGACAAA 120

```
||||| ||||||| || ||||||||||||| | | ||||||| | | |||
Db 149 AGAAAAACCTGACCCAGATGTTTATATAAGGAAGGAACAGAAAGAGGGAACGAAA 208
QY 121 AGAAAGGAAATAAGA 137
Db 209 AGAAAGGAGAAATAAAA 225

RESULT 6
LOCUS AI890775 547 bp mRNA linear EST 07-MAR-2000
DEFINITION wm95fl1.x1 NCI_CGAP_Ut2 Homo sapiens cDNA clone IMAGE:2443725 3'
similar to SW:CHD1_HUMAN O14646 CHROMODOMAIN-HELICASE-DNA-BINDING
PROTEIN 1 ; mRNA sequence.
ACCESSION AI890775
VERSION AI890775.1 GI:5595939
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 547)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: Greg Lennon, Ph.D.
Clone Distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/dbp/image/image.html
Insert Length: 1924 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 418.
Location/Qualifiers
1..547
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="IMAGE:2443725"
/clone_lib="NCI_CGAP_Ut2"
/tissue_type="moderately-differentiated endometrial
adenocarcinoma 3 pooled tumors"
/lab_host="DH10B"
/note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.85 kb. Life Technologies catalog #:
11539-012"
BASE COUNT 114 a 118 c 85 g 230 t
ORIGIN

Query Match 59.1%; Score 81; DB 9; Length 547;
Best Local Similarity 80.5%; Pred. No. 0.086;
Matches 107; Conservative 0; Mismatches 25; Indels 1; Gaps 1;

QY 1 GATGGATTGTTTCAGTGAACATCCACATAAAAAATAAAAGCAGAAAAAGAAATGA 60
Db 132 GATGAGATCAGTTCTGTGAACATCCAAATAAAAAATAAAACAGAAAAG-AGACAGTGA 74

QY 61 AGAAAAAGATGACCCAGATGTTGTTATAAGAGGAGCTGGAGAAAAAGAGACACAA 120
Db 73 AGAAAAACCTGACCCAGATGTTTATATAAGGAAGGAACCAAGAAAGAGGGAACGAAA 14

QY 121 AGAAAAAGGAAAT 133
Db 13 AGAAAGGAGAAAT 1
```

```
RESULT 7
LOCUS AW387264 249 bp mRNA linear EST 04-FEB-2000
DEFINITION MR1-ST0088-101199-003-b09 ST0088 Homo sapiens cDNA, mRNA sequence.
ACCESSION AW387264
VERSION AW387264.1 GI:6891923
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 249)
AUTHORS HCGP http://www.ludwig.org.br/ORESTES.
TITLE The FAPESP/LICR Human Cancer Genome Project
JOURNAL Unpublished (1999)
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?t1=MR1&t2=MR1-ST0088-
101199-003-b09&t3=1999-11-10&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 27
High quality sequence stop: 248.
Location/Qualifiers
1..249
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="ST0088"
/dev_stage="Adult"
/note="Organ: stomach; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT 52 a 52 c 41 g 104 t
ORIGIN

Query Match 53.9%; Score 73.8; DB 9; Length 249;
Best Local Similarity 75.9%; Pred. No. 0.95;
Matches 104; Conservative 0; Mismatches 32; Indels 1; Gaps 1;

QY 1 GATGGATTGTTTCAGTGAACATCCACATAAAAAATAAAAGCAGAAAAAGAAATGA 60
Db 159 GATGAGATCAGTTCTGTGAACATCCAAATAAAAAATAAAACAGAAAAG-AGACAGTGA 101

QY 61 AGAAAAAGATGACCCAGATGTTGTTATAAGAGGAGCTGGAGAAAAAGAGACACAA 120
Db 100 AGAAAAACCTGACCCAGATGTTTATATAAGGAAGGAACCAAGAGGAGGGAAGCACA 41

QY 121 AGAAAGGAAATAAGA 137
Db 40 AGCAAGAGAGAACTCAA 24

RESULT 8
LOCUS BE906261 782 bp mRNA linear EST 20-OCT-2000
DEFINITION 601502247F1 NIH_MGC_70 Homo sapiens cDNA clone IMAGE:3904306 5',
mRNA sequence.
ACCESSION BE906261
VERSION BE906261.1 GI:10399603
KEYWORDS EST.
SOURCE human.
```

```

ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 782)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM9709 row: m column: 11
High quality sequence stop: 35.
Location/Qualifiers
1. 782
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NIH-MGC_70"
/tissue_type="epitheloid carcinoma"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: pancreas; Vector: pCMV-SPORT6; Site: 1; Not I;
Site: 2; SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.1 kb. Library constructed by Life
Technologies."
BASE COUNT      530 a      64 c      165 g      23 t
ORIGIN

Query Match      42.5%; Score 58.2; DB 10; Length 782;
Best Local Similarity 68.1%; Pred. No. 47;
Matches 81; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY 19 AACATCCACATATAAAGCAGCAAAAGAAATGAAGAAAAGATGAGCAAGATGAGCCAGA 78
Db 346 AAAAAACACAAAAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 405
QY 79 GATTGTATTAAGAGGAGGAGCTGAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 137
Db 406 GAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 464

RESULT 9
BM415214/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchinea;
Tylenchoidea; Heteroderidae; Heteroderinae; Globodera.
1 (bases 1 to 973)
Heer, J., Sosinski, B., Pokrzywa, R., Wary, A. and Opperman, C.
Mixed Stage EST's from Globodera pallida, the potato cyst nematode
Unpublished (2001)
Contact: Opperman, C
Center for the Biology of Nematode Parasitism
NC State University; IACR-Rothamsted
Campus Box 7616; Raleigh, NC 27695, USA
Tel: 919.515.6699
Fax: 919.515.9500
Email: warthog@unity.ncsu.edu
GT11-12PCN_R_H11-GT11-12_R_092.ab1.
Location/Qualifiers
1. 973

```

```

/organism="Globodera pallida"
/db_xref="taxon:36090"
/clone_lib="Mixed Stage EST's from Globodera pallida, the
potato cyst nematode"
/notes="Vector: lambda GT11; This is a collaborative effort
between IACR-Rothamsted and North Carolina State
University. The library was constructed from mixed stage
G. pallida in lambda GT11 by Paul Burroughs,
IACR-Rothamsted."
BASE COUNT      231 a      12 g      404 t      45 others
ORIGIN

Query Match      42.2%; Score 57.8; DB 10; Length 973;
Best Local Similarity 70.6%; Pred. No. 48;
Matches 77; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 27 ACATATAAATAAAGCAGCAAAAGAAATGAAGAAAAGATGAGCCAGATGTGTA 86
Db 935 AGAGAAAAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 876
QY 87 TAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 135
Db 875 AAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 827

RESULT 10
BG548029
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 200)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: CLONTECH Laboratories, Inc.
CDNA Library Preparation: CLONTECH Laboratories, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1543 row: e column: 12
High quality sequence stop: 1.
Location/Qualifiers
1. 200
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NIH-MGC_77"
/lab_host="DH10B (T1 phage-resistant)"
/notes="Organ: lung; Vector: pDNR-LIB (Clontech); Site: 1;
Site: 2; SfiI (ggccattatggcc); Site: 3; SfiI (ggccattatggcc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
5'-ATTCTAGAGCGGAGCGGCGGAGATG-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.9
kb (range 0.5-4.0 kb). 12/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA). Note: this is a NIH-MGC Library."
BASE COUNT      139 a      9 c      47 g      5 t
ORIGIN

```







```
fly), genomic survey sequence.
```

```
ACCESSION      AL064007.1   GI:494187
VERSION        AL064007.1
KEYWORDS       GSS.
SOURCE         fruit fly.
ORGANISM       Drosophila melanogaster
               Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
               Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
               Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE      1 (bases 1 to 338)
AUTHORS        Genoscope.
TITLE          Direct Submission
JOURNAL        Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
               - Web : www.genoscope.cns.fr)
COMMENT        Determination of this BAC-end sequence was carried out as part of a
               collaboration with the Berkeley Drosophila Genome Project (BDGP).
               The BDGP is constructing a physical map of the Drosophila
               melanogaster genome using these BACs. For further information
               please see http://www.fruitfly.org The BDGP Drosophila
               melanogaster BAC library was prepared by Kazutoyo Osoegawa and
               Aaron Mammoser in Pierer de Jong's laboratory in the Department of
               Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
               NY. The library is named RPCI-98 and was constructed by partial
               EcoRI digestion of Drosophila DNA provided by the BDGP from the
               isogenic strain Y2; cn bw sp, the same strain used for the BDGP's
               Pl and EST libraries. A more detailed description of the library
               and how to order individual BAC clones, the entire library, or
               filters for hybridization from the BACPAC Resource Center can be
               found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
```

```
FEATURES             Location/Qualifiers
     source            1..338
                       /organism="Drosophila melanogaster"
                       /db_xref="taxon:7227"
                       /clone_lib="RPCI-98"
                       /clone="BACR13M18"
                       /note="end : T7"
BASE COUNT           5 a    13 c    6 g    294 t    20 others
ORIGIN
```

```
1..338
+=====+
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPCI-98"
/clone="BACR13M18"
/note="end : T7"
```

```
Query Match      40.1%; Score 55; DB 12; Length 338;
Best Local Similarity 58.8%; Pred. No. 1.7e+02;
Matches 70; Conservative 12; Mismatches 37; Indels 0; Gaps 0;
```

```
QY 19 AAACATCCACATAAAAAAATAAGACGAGAAAAGAAATGAAGAAAAAGATGAGGCAGA 78
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 189 AAAAAAAAAAAAAAAAAAATMAAAMWMAAAAAAAAAAAAAAAAAAGARRRAAAAAAAAA 130
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 79 GATTGTATATAAGAACGACTGGAGAAAAAAGAGAGACAAAAAGAAAGAAATAAGA 137
|| || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 129 AAAAAAAAAAAAAAAAAAARAAAAAAAAAAAAAAAAARAAAAAAAAAGAAAWAAGA 71
|| || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

```
Search completed: August 2, 2002, 16:53:19
Job time: 12280 sec
```

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 2, 2002, 20:53:02 ; Search time 1205.34 Seconds  
(without alignments)  
1874.541 Million cell updates/sec

Title: US-08-973-363-15  
Perfect score: 1316  
Sequence: 1 ATTATCGCTACTACAAA.....AATAAGCAACAAGGCTGC 1316

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_032802.\*  
1: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1303.6	99.1	1316	18 AAT42754	Chicken CHD-W gene
2	1104.6	83.9	6608	18 AAT42751	Chicken CHD-1A gen
3	193.2	14.7	4222	23 AAS85967	DNA encoding novel
4	157.2	11.9	6240	23 ABL06443	Drosophila melanog
5	157.2	11.9	9933	23 ABL06442	Drosophila melanog
6	142.6	10.8	1311	19 AAV59280	Altered telomere r
7	122.8	9.3	474	22 AAL36123	Human breast cance
8	122.8	9.3	862	22 AAL18193	Human breast cance
9	100.4	7.6	153	18 AAT42758	Chick CHD-W gene f

10	92.4	7.0	153	18 AAT42757	Chick CHD-1A gene
11	87.6	6.7	153	18 AAT42759	Great tit CHD-W ge
12	87.4	6.3	249	21 AAO1247	Human colon cancer
13	67.4	5.1	153	18 AAT42756	Mouse CHD-1 gene (
14	65.2	5.0	421	22 AAK88882	Human digestive sy
15	65.2	5.0	421	22 AAI57603	Human colorectal c
16	63.4	4.8	297	22 AAL08378	Human breast cance
17	53	4.0	3399	17 AAT05868	Chicken leucocytoz
18	52.4	4.0	1686	16 AAO87587	DNA encoding Leuco
19	52.4	4.0	4590	7 AAN60472	Sequence encoding
20	49.2	3.7	423	22 ABA48060	Human breast cell
21	49.2	3.7	423	22 ABA65943	Human foetal liver
22	49.2	3.7	423	22 ABA33035	Probe #11501 for g
23	49.2	3.7	423	22 AAK14364	Human brain expres
24	49.2	3.7	423	22 AAK40093	Human bone marrow
25	49.2	3.7	423	22 AAI20873	Probe #10806 for g
26	49.2	3.7	423	22 AAI46111	Probe #14797 used
27	49.2	3.7	423	22 AAI06583	Probe #6574 used t
28	47	3.6	8576	24 ABL34229	Human immune syste
29	46.4	3.5	487	22 ABA42932	Human breast cell
30	46.4	3.5	487	22 ABA53351	Human foetal liver
31	46.4	3.5	487	22 ABA23131	Probe #1597 for ge
32	46.4	3.5	487	22 AAK01618	Human brain expres
33	46.4	3.5	487	22 AAK27061	Human bone marrow
34	46.4	3.5	487	22 AAI11660	Probe #1193 for ge
35	46.4	3.5	487	22 AAI32959	Probe #1645 used t
36	46.4	3.5	487	22 AAI01586	Probe #1577 used t
37	44.4	3.4	3095	11 AAO03875	Sequence encoding
38	43	3.3	1998	21 AAO70212	Plasmodium falcipa
39	42.8	3.3	389	5 AAN40273	Sequence of the fl
40	42.6	3.2	597	22 AAT78832	Degenerate polynuc
41	41.6	3.2	4766	15 AAO70102	Malarial PfEMP3 ep
42	41.6	3.2	7326	21 AAO70095	Plasmodium falcipa
43	41.6	3.2	42299	22 AAK68932	Human immune/haema
44	41.4	3.1	427	20 AAK61370	DNA encoding a hum
45	41.4	3.1	708	23 AAS69547	DNA encoding novel

ALIGNMENTS

RESULT 1  
ID AAT42754 standard; cDNA; 1316 BP.  
AC AAT42754;  
XX  
XX 12-MAR-1997 (first entry)  
XX  
XX Chicken CHD-W gene (partial sequence).  
XX  
XX Bird; sex determination; chromodomain-Helicase-DNA binding 1 Avian;  
XX CHD-W; chromodomain-Helicase-DNA binding on the W chromosome; ss.  
XX  
XX Gallus sp.  
XX  
XX WO9639505-A1.  
XX  
XX 12-DEC-1996.  
XX  
XX 05-JUN-1996; 96WO-GB01341.  
XX  
XX 06-JUN-1995; 95GB-0011439.  
XX  
XX (ISIS-) ISIS INNOVATION LTD.  
XX  
XX Griffiths R, Tiwari B;  
XX  
XX WPI; 1997-043127/04.  
XX  
XX Avian chromodomain-helicase-DNA binding genes determine sex in  
XX birds - used for sex determ. and to control sex of progeny

PS Claim 1; Fig 8; 76pp; English.

XX The chicken CHD-W gene (AAT42754) acting alone or in conjunction with the closely related CHD-1A gene (AAT42751) is suggested to initiate female development in birds. The sequence of CHD-W was deduced from 2 clones isolated from a 10-day chick embryo library using a fragment of the CHD-1A gene as a probe. The CHD-W gene is located on the W chromosome. Probes based on CHD-W and CHD-1A give a W chromosome-specific signal on hybridisation to genomic DNA of a non-ramite bird and can be used for sex determination of a bird. CHD-W nucleic acids can also be used to control the sex of progeny of a bird.

XX Sequence 1316 BP; 492 A; 208 C; 306 G; 304 T; 6 other;

SQ

Query Match 99.1%; Score 1303.6; DB 18; Length 1316;  
Best Local Similarity 99.7%; Pred. No. 0;  
Matches 1312; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ATTATCGGCTAGTCACAAAAGGATCAGTAGAAGAGATATTCTTGAAAGAGCCAGAGAA 60  
|||||  
Db 1 attatcggctagtcacaaaagatcagtagaagaatatcttgaagaagcccaagaaa 60  
|||||

QY 61 AAGATGGTGTATGATCATTTAGTATTCAGAGAAATGCACACAGGAGAAACTGTACTA 120  
|||||  
Db 61 aagatggtgtatgatactttagtattcagatgagaaatgcacacaggggaaactgtacta 120  
|||||

QY 121 CATACAGGCTCTACTCTTCACAGCTCAACACCTTTTAAAGGAAGAGTTATCAGCAATT 180  
|||||  
Db 121 catacaggtctactcttcaagctcaacacaccttttaagaagaaggttatcagcaatt 180  
|||||

QY 181 TTGAAGTTTGGTCTGAGGAACCTTTTAAAGAACCTTGAANNNGAAGAGGAGCCTCAG 240  
|||||  
Db 181 ttgaagtgttggctgaggaacttttaagaagaaccttgaannngaagaagagagcctcag 240  
|||||

QY 241 GAGATGGATATAGATGAATCTCTGAGAGNCTGAACTCGAGAAATAGTCAGGCCCA 300  
|||||  
Db 241 gagatggatatagatgaatctctgagagngctgaaactcgagaaatgagtcaggccca 300  
|||||

QY 301 TTAACGTAGGAGATGAGTTACTTTTCACAGTTCAAGGTAGCTTAACCTTTTCCATATGGAT 360  
|||||  
Db 301 ttaactgtaggagatgagttacttttcacagttcaaggttagcttaaccttttccaatatggat 360  
|||||

QY 361 GAAGATGACATTTGAATGGACACAGAACAAATCTAAGAACTGGGAAGAAATCATTTCCA 420  
|||||  
Db 361 gaagatgacatggaatggacacagaaacaaatctaagaacctggaannngaagaagagcctcag 420  
|||||

QY 421 GAAGTTTCAGTGGCGAGAAATAGAGGGNNGAAGACAAAAGAACTTGAAGAAATATAT 480  
|||||  
Db 421 gaagtttcagtggcgagaaatagagggngngaagacaaaagaaacttgaagaaatataat 480  
|||||

QY 481 ATGCTTCCAAAGATGAGAACTCTGCAAAACAGATCAGCTTTAATGGAATGAAGGGAGA 540  
|||||  
Db 481 atgcttccaaagatgagaaactctgcaaaacagatcagcttttaatggaaatgaaggaga 540  
|||||

QY 541 TGCAGTAGGAGCAGAGATATTCTTGATCTGATAGTATTCATCTCAGAAAAGAAACCA 600  
|||||  
Db 541 tgcagtaggagcagaagatatcttgatctgatatgattccatctcagaagaagaaacga 600  
|||||

QY 601 CCAAAAAACGTGGAGCAGCAGAACTATCCCGGTGAAAAATTAAGGATTTAGTGAT 660  
|||||  
Db 601 ccaaaaaacgtggagcagcagaaactatcccggtcaaaacattaaaggatttagtgat 660  
|||||

QY 661 GCAGAGATTAGACGATTTATCAAGAGTTACAGAAATTTGGTGGCCAGTTCAAGGTTA 720  
|||||  
Db 661 gcagagattagacgatttatcaagagtttacagaaatttgggtggccagttgaaagggtta 720  
|||||

QY 721 GATGCTATAGCTAGAGATGCTGAGTGTGATAATCTGAAACACAGACCTTAGACGCTCG 780  
|||||  
Db 721 gatgctatagctagagatgctgagtagttgataaatctgaaacagacaccttagacgctcg 780  
|||||

QY 781 GGAGAACTTGTCATTAATGGATGCATTAAGGCTTTAAATGATTAATGACGTTTGGTCAAGGA 840  
|||||

Db 781 ggagaactgtacataatggtgcatgaaggttttaataatgaataagactttggtcaagga 840  
|||||

QY 841 AGAACAGGTGGTAGATTTCGGGAAAGTTAAAGGCCCAACATTCGGAATACAGAGTGCAG 900  
|||||

Db 841 agaacaggtggtagatttggaaagttaaaggcccaacattccgaatagcagagtgag 900  
|||||

QY 901 GTGAATGCAAGCTAGTCTATCTTCACGAAGAAGAGTTGGCAGCATTTGATTAATCGATT 960  
|||||

Db 901 gtgaatgcaagctagctcattcttcacgaagaagagttggcaccattgataaatacgatt 960  
|||||

QY 961 CTTTCAGATCCAGAAAGAAAGATATGTCATCCCATACACACCAAGAGCAGTCTAT 1020  
|||||

Db 961 ctttcagatccagaagaagaaagatatgtcatcccataccacacccaagcagctcat 1020  
|||||

QY 1021 TTTGATATAGATTGGGTAAAGAGATGATTCCAATCTGTTAATAGGCATCTATGAATAT 1080  
|||||

Db 1021 ttgatataagattgggttaaagaagatgattccaatctgttaataggtcatctatgaat 1080  
|||||

QY 1081 GGTATGCGCAGTTGGGAAATGATAAAATGGATCTCTGATCTCAGTTTGACACAGAAAGATT 1140  
|||||

Db 1081 ggttatggcagttggaaatgataaaaatggatcctgactcagtttgacacagaagatt 1140  
|||||

QY 1141 TTACCTGATGATCCAGATAAGAAACCCAGGCTTAAGCAGTTACAGACTCGTGAGATTAC 1200  
|||||

Db 1141 ttacctgatgatccagataaagaacccaggtctaaagcagttacagactcgtgcagattac 1200  
|||||

QY 1201 CTCATTAAATTAAGTAAGAACCTTGCAAGAAAGAGAGCAGACAGACTTCTGCTGTCGA 1260  
|||||

Db 1201 ctcattaaattactgaataaagaccttgcaagaagaagagcagagacttctgctgtgca 1260  
|||||

QY 1261 GGCAATTTCAAGAGGAGAGAAACACAGAACTAAGAAATAAAGCAACAAAGGCTGC 1316  
|||||

Db 1261 ggcaatttcaagaggagagaaaaaacagaagttaagaagaataaagcaacaaggctgc 1316  
|||||

RESULT 2  
AAT42751  
ID AAT42751 standard; cDNA; 6608 BP.  
XX  
AC AAT42751;  
XX  
XX 12-MAR-1997 (first entry)  
DT  
XX Chicken CHD-1A gene.  
DE  
XX Bird; sex determination; chromodomain-Helicase-DNA binding 1 Avian;  
KW CHD-W; chromodomain-Helicase-DNA binding on the W chromosome; ss.  
XX  
OS Gallus sp.  
XX  
FH Key Location/Qualifiers  
CDS 228..5390  
FT /\*tag= a  
XX  
PN WO9639505-A1.  
XX  
PD 12-DEC-1996.  
XX  
PF 05-JUN-1996; 96WO-GB01341.  
XX  
PR 06-JUN-1995; 95GB-0011439.  
XX  
PA (ISIS-) ISIS INNOVATION LTD.  
XX  
PI Griffiths R, Tiwari B;  
XX  
DR WPI; 1997-043127/04.  
XX  
PT Avian chromodomain-helicase-DNA binding genes determine sex in  
PT birds - used for sex determ. and to control sex of progeny  
XX  
PS Claim 1; Fig 5; 76pp; English.

XX CC The chicken CHD-W gene (AA142754) acting alone or in conjunction with  
CC the closely related CHD-1A gene (AA142751) is suggested to initiate  
CC female development in birds. The sequence of CHD-1A was deduced  
CC from 3 clones isolated from a stage 10-12 chicken embryo cDNA  
CC library using a great tit CHD-W sequence (see also AA142755) as probe.  
CC The CHD-1A (A - Avian) gene shows close identity to the mouse CHD-1  
CC gene (see also AA142756-57). It is located on an autosome or X  
CC chromosome. Probes based on CHD-W and CHD-1A give a w chromosome-  
CC specific signal on hybridisation to genomic DNA of a non-rare  
CC bird and can be used for sex determin. of a bird. CHD-1A nucleic  
CC acids can also be used to control the sex of the progeny of a bird.  
XX SQ Sequence 6608 Bp; 2289 A; 1207 C; 1459 G; 1653 T; 0 other;

Query Match 83.9%; Score 1104.6; DB 18; Length 6608;  
Best Local Similarity 90.4%; Pred. No. 2.3e-279;  
Matches 1176; Conservative 0; Mismatches 125; Indels 0; Gaps 0;  
QY 1 ATTTATCGGCTAGTCACAAAGGATCAGTAGAAGAGATATTCTTGAAGAGCCAGAAA 60  
DB 2943 atttattcggttagtcacaaaagatcagtagaagaagattcttgaagagccaaagag 3002  
QY 61 AAGATGGTGTAGATCAATTTAGTATTCAGAGATGCACACACAGGAAACTGTACTA 120  
DB 3003 aagaatggttagaccatttagtaattcagagatggaacacagagaaaactgtctcg 3062  
QY 121 CATACAGGCTCTACTCTCAAGCTCAACACCTTTTAATGAAGAGAGATTATCAGCAATT 180  
DB 3063 caacaggtctcaactccatccatcagctcacaccttttaataaagaagagttcagctatt 3122  
QY 181 TTGAAGTTGGTGTGAGGAACCTTTTAAAGAACCTTGAANNNGAAGAGAGGAGCTCAG 240  
DB 3123 ttgaagttggtgtgaggaactctttaaagaacctgaaggaagacagagagcccaag 3182  
QY 241 GAGATGGATATAGATGAATTTGGAACAGACAAATCTAAGAACTGGGAAGAAATCATTTCCA 300  
DB 3183 gaattggtatagatgaattcttgaagagagctgaacactcggaataatgagccaggtcca 3242  
QY 301 TTAACCTAGGAGATGAGTTACTTTACAGTTCAGAGTCAAGTTCATCTTCCAAATATGAT 360  
DB 3243 ttgactgtagggatgagttgctttcacagttcaaggtggtgcaacttttcccaattgagat 3302  
QY 361 GAAGATGACATTGAATTTGGAACAGACAAATCTAAGAACTGGGAAGAAATCATTTCCA 420  
DB 3303 gaagatgattgagttggaacacagaagaattcaagaattggaagaatcatccca 3362  
QY 421 GAAGTTTCAGTGGCAGCAGATAGAGGGNNGAAGACAAAAGAACTTGAAGAAATATAT 480  
DB 3363 gaatcccaacggaagatagagagaggaagaaacaaaagaacttgaaataatatac 3422  
QY 481 ATGCTTCCAGAAATGAGAACTTCGAAACAGATCAGCTTTAATGGAATGAAGGAGA 540  
DB 3423 atgctccgagatgagaactgtgcaaacacagatcagcttttaactgaggaaggaaga 3482  
QY 541 TGCATGAGGACACAGATATTTCTGGATCTGATGATTCCTCTCAGTCAAGAAAGAA 600  
DB 3483 cgcagtaggagcagaagatattctggaactgtagtgcactccatcacagaagaagaacgg 3542  
QY 601 CCAAAAACGTTGACGACCACTATTCCTCCGTTGAACAACTTAAAGGATTTAGTAT 660  
DB 3543 ccaaaaagcgtggaagacctgcacacctctctcgagaaaattattaaagatttagttagt 3602  
QY 661 GCAGAGATGACGATTTATCAAGAGTTTACAAGAAATTTGGTGCCCGCAGTTGAAGGTTA 720  
DB 3603 gcagagatcagcgcgtttatcaagagattacaagaataatttgggtgcccctctggaagggtta 3662  
QY 721 GATCCTATAGCTAGAGTGTGAGCTAGTGTGATAAATCTGAACACAGCTTACAGCTGTG 780  
DB 3663 gatcgtgtagctagatgctgaactggtgataaactctgacagacagaccttagagtttg 3722  
QY 781 GGAGAACTTGTCATATAATGATGCATTAAGGCTTTAAATGATATGACTTTGTCACAGA 840

DB 3723 ggtgaactgtacataatgtagatgcattaaaggctttaaaggcaaatctctggaagaag 3782  
QY 841 AGAACAGGTTGGTAGATTGGGAAAGTTAAAGGCCCAACATTCGGAATAGCAGGAGTCCAG 900  
DB 3783 agagcagaggttagactgtggaaagttaaagcccaacgtttcgaatctcagggagtcag 3842  
QY 901 GTGAATGCAAGCTAGTACTTCTCAGCAAGAAGAGTTGGCACCATTTCATAAATCGATT 960  
DB 3843 gtgaatgcaaaactagtcattctctcaggaagagctggtgcaccactgcacaaatccatt 3902  
QY 961 CTTTCAGATCCAGAAAGAAAGAAATATGTCTATCCCATACACACCAAGAGAGCTCAT 1020  
DB 3903 ctttcagatccagaagaagaaagatgattgcatcccatgccacacaaaggtctgctcac 3962  
QY 1021 TTTGATATAGATTGGGTAAAGAGATGATTCCAATCTGTTAATAGGCACTATGATATAT 1080  
DB 3963 ttgatatagattgggttaaagaagatgattccaatctgttagtaggcatctatgaatat 4022  
QY 1081 GGTATGGCAGTTGGGAAATGATTAATAAATGGATCTCTCAGTTTGACACAGAGATT 1140  
DB 4023 ggtatgagcagctgggaagtataaaatggatccagatctcagcttaacacagaagatt 4082  
QY 1141 TTACCTGATGATCCAGATAGAAACCCAGGCTAAGCAGTTACAGACTCTGTCGAGATTAC 1200  
DB 4083 ttacctgtagaaccagacaagaacccccaggaagcagctacagaccgctgcagactac 4142  
QY 1201 CTCATTAAATTTACTGTAATGAAGACCTTTGCAAGAAAGAACGACACAGAGACTTTGCTGTGCA 1260  
DB 4143 ctcaataaattactgaataaagaccttgcagaagaaggaacacaaaggcttgcgtgtagca 4202  
QY 1261 GGCAATTTCAAGAGGAGAGAAACAAAGAGTAAAGAGAAATAA 1301  
DB 4203 ggcattccaaagagaaggaagacagaataagaagaataa 4243

RESULT 3  
ID AA585967 standard; cDNA: 4222 BP.  
XX AC AA585967;  
XX DT 13-FEB-2002 (first entry)  
XX DE DNA encoding novel human diagnostic protein #21771.  
XX KW Human: chromosome mapping; gene mapping; gene therapy; forensic;  
XX KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX OS Homo sapiens.  
XX PN WO200175067-A2.  
XX PD 11-OCT-2001.  
XX PF 30-MAR-2001; 2001WO-US08631.  
XX PR 31-MAR-2000; 2000US-0540217.  
XX PR 23-AUG-2000; 2000US-0649167.  
XX PA (HYSE-) HYSEQ INC.  
XX PI Drmanac RT, Liu C, Tang YT;  
XX DR WPI: 2001-639362/73.  
XX DR P-PSDB: ABG21780.  
XX PT New isolated polynucleotide and encoded polypeptides, useful in  
XX PT diagnostics, forensics, gene mapping, identification of mutations  
XX PT responsible for genetic disorders or other traits and to assess  
XX PT biodiversity -  
PS Claim 1; SEQ ID No 21771; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AA594564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 4222 BP; 1040 A; 1041 C; 883 G; 1258 T; 0 other;

Query Match 14.7%; Score 193.2; DB 23; Length 4222;  
Best Local Similarity 82.2%; Pred. No. 1.8e-40;  
Matches 222; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

QY 714 AAGGTTAGTGTATAGCTAGAGATGCTGAGCTAGTGTGATAAATCTGAACAGACCTTAG 773  
II IIIIIII III IIIIIIIII IIIIIIIII IIIIIIIII IIIIIIIII IIIIIIIII  
Db 2753 AAAATTAGATGCAATTGCTCGAGTGTCTGAGTTAGTGTGATAAGTCAGAAACAGCTTAG 2694

QY 774 AGCTCTGGGAGAACTGTACATAATGGATGCAATTAAGGCTTTAAATGATAATGACTTTGG 833  
II IIIIIIIII III IIIIIIIII IIIIIIIII IIIIIIIII IIIIIIIII IIIIIIIII  
Db 2693 ACAGCTGGGAGAAATGGTATGATTAATGGTTGCAATTAAGCATTTAAAGGATGTTCTTCAGG 2634

QY 834 TCAAGGAAGAACAGGTGGTGTAGATTTGGGAAAGTTAAAGGCCCAACATTCGGAATAGCAGG 893  
II III IIIIIIIII IIIIIIIII III IIIIIII IIIIIIIII IIIIIIIII IIIIIIIII  
Db 2633 AACAGAACGACAGGTGGTGTAGACTCGGAAAGTGAAGGTCACACATTCGGAATATCAGG 2574

QY 894 AGTCAGGTGAATGCAAGCTAGTCAATTTCTCAGGAAGAGGTGGCACCATTCATAA 953  
II IIIIIIIII III IIIIIIIII III IIIIIIIII III IIIIIIIII III IIIIIIIII  
Db 2573 AGTACAGGTGAATGCGCAACTAGTCACTCCCATGAAGAAGTAATTAATACCTTTGCACAA 2514

QY 954 ATCGATTCTCTCAGATCCAGAGAAGGAA 983  
II III IIIIIIIII IIIIIIIIIII III IIIIIIIII III IIIIIIIII III IIIIIIIII  
Db 2513 ATCCATTCTCTGTATCCAGAAAGAA 2484

RESULT 4  
ABL06443  
ID ABL06443 standard; cDNA; 6240 BP.  
XX  
AC ABL06443;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 13811.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical; gene; ss.  
XX  
OS Drosophila melanogaster.  
XX  
PN WO200171042-A2.  
XX  
XX  
PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US09231.  
XX  
XX 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.  
XX  
PA (PEKE ) PE CORP NY.  
XX  
PI Venter JC, Adams M, Li PWD, Myers EW;  
XX  
DR WPI; 2001-656860/75.  
DR P-PSDB; ABB62340.  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX  
PS Claim 1; SEQ ID NO 13811; 21pp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins  
CC (ABB57737-ABB72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 6240 BP; 1675 A; 1682 C; 1760 G; 1123 T; 0 other;

Query Match 11.9%; Score 157.2; DB 23; Length 6240;  
Best Local Similarity 49.9%; Pred. No. 5.8e-31;  
Matches 652; Conservative 0; Mismatches 564; Indels 90; Gaps 7;

QY 1 ATTTATCGCTAGTCACAAAAGCATCAGTAGAAGAGATATTTCTTGAAGAGACCCAGAAA 60  
Db 2978 atttttcgctgtgaccgccggtcggtggaggaaacagatcgtggagcgagccaaacag 3037

QY 61 AAGATGGTGTAGATCATTTAGTGTATTCAGAGAATGGACACACACAGGAAAATCTGTACTA 120  
Db 3038 aaaaatggtagtaccacctggctcattcagcggatggacactacagagcagcgtgtctg 3097

QY 121 CATAAGGCTCTA-----CTCCTTTCAAGCTCAACACCTTTTAAATGAAGAGAGTTATCA 174  
Db 3098 gataagagcggaacggcgcaactcgtctaactcgaatccgttcaacagagcagatcttct 3157

QY 175 GCAATTTGAAGCTTTGGTCTGAGGAACCTTTTAAAGAACCTCAANNNGAAGAGAGAG 234  
Db 3158 gctatcctcgaagtttggtagcagagagcgtgttcaagga---cgaacagagcagcagcag 3214

QY 235 CCTCAGGAGATGGATATAGTAAATCTCTGAAGAGGNCCTGAACCTCGAGAAAATGAGTCA 294  
Db 3215 gacttggtttgtacatgatgaaattctgcagggcgagagcccgcaacga-----a 3268

QY 295 GCCCATTAACGTGAGGAGATGAGTTACTTTACAGTTCACAGTTCACAGTTCACATTTTCCAAT 354  
Db 3269 gatccggaatgccagcagcagcactgtgttcgccttcaagtgccagcagcagcagcagcag 3328

QY 355 ATGCGATGAAGATGACATTGAATTTGAACCAACAGAAAATCTA----- 396  
Db 3329 ttccgaggggaaccgcagcagcgttcggttagcgaagcagcgaacaaatgccccgcgcagggag 3388

QY 397 -----AGAAACTGGGAAGAAATCATTTCCAGAAGTTTCAGTGGCGAGCAATAGAGGGG 447  
Db 3389 gatgacagtaaggactggagacatcattccggaagcgttcgtaaaagcaatcgacgat 3448

QY 448 NNGGAAAGCAAAACAACTTGAAGAAATATATATGCTTTCCAGAAATGAGAACTGTGCA 507  
Db 3449 caggagcgcaaaagagatggaagatcttacttgccaccgccgaggaacagcgcgct 3508

QY 508 AAACAGATCAGCTTTAATGGAATGAAGGAGATCAGCTAGGAG----- 551  
Db 3509 aaccacaaacgaagggaagcgtgtgagcgtggaaggggggcaagggaacacagcagcagcagcag 3568







```
XX PF 10-JAN-2001; 2001WO-US00798.
XX PR 14-JAN-2000; 2000US-0176077.
XX PR 14-MAR-2000; 2000US-0189167.
XX PR 24-MAR-2000; 2000US-0192099.
XX PR 29-MAR-2000; 2000US-0193480.
XX PR 15-MAY-2000; 2000US-0205230.
XX PR 09-JUN-2000; 2000US-0211315.
XX PR 15-JUL-2000; 2000US-0220534.
XX PR 25-JUL-2000; 2000US-0220534.
XX PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX PI Lillie J, Xu Y, Wang Y, Steinmann K;
XX DR WPI; 2001-451856/48.
XX PT New peptide useful as a marker for the diagnosis of breast cancer -
XX PS Claim 1; Page 3424; 3695pp; English.
XX CC The invention relates to human breast cancer expressed polynucleotides
XX CC (AAL07544-AAL26789) and methods of assessing whether a patient is
XX CC afflicted with breast cancer by examining the correlation between the
XX CC expression of certain markers and the cancerous state of breast cells.
XX CC The polynucleotides and encoded polypeptides are potential markers for
XX CC detecting, diagnosing, monitoring, characterising treating and
XX CC potentially preventing breast cancer. The polynucleotides and encoded
XX CC polypeptides are also useful for isolating compounds with cytostatic
XX CC activity.
XX SQ Sequence 474 BP; 156 A; 72 C; 83 G; 163 T; 0 other;

Query Match          9.3%; Score 122.8; DB 22; Length 474;
Best Local Similarity 86.1%; Pred. No. 2.2e-22;
Matches 136; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

QY 985 AGATATGTCATCCATACACACCAAGCAGCTCATTTGATATAGATTGGGTAAGAA 1044
DB 307 aggtatactatccatgcacaaaggcagctcatttgatagactggggcaagaa 366
QY 1045 GATGATTCATCTGTTAATAGGCATCTATGATATGTTATGGGTAAGGATA 1104
DB 367 gatgctcccaattgcttaattggcatctatgaatggatggatgggaagtggaaatgatt 426
QY 1105 AAAATGGATCTGATCTCAGTTTGACACAGAAGATTTT 1142
DB 427 aaatggatcctgacctcagctctaacaacacaaaggatt 464

RESULT 8
AAL18193
ID AAL18193 standard; cDNA; 862 BP.
XX AC AAL18193;
XX DT 07-DEC-2001 (first entry)
XX DE Human breast cancer expressed polynucleotide 10650.
XX KW Human; breast cancer; cell marker; cytostatic; ss.
XX OS Homo sapiens.
XX PN W0200151628-A2.
XX PD 19-JUL-2001.
XX PF 10-JAN-2001; 2001WO-US00798.
XX PR 14-JAN-2000; 2000US-0176077.
XX PR 14-MAR-2000; 2000US-0189167.
```

```
XX PR 24-MAR-2000; 2000US-0192099.
XX PR 29-MAR-2000; 2000US-0193480.
XX PR 15-MAY-2000; 2000US-0205230.
XX PR 09-JUN-2000; 2000US-0211315.
XX PR 25-JUL-2000; 2000US-0220534.
XX PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX PI Lillie J, Xu Y, Wang Y, Steinmann K;
XX DR WPI; 2001-451856/48.
XX PT New peptide useful as a marker for the diagnosis of breast cancer -
XX PS Claim 1; Page 1901; 3695pp; English.
XX CC The invention relates to human breast cancer expressed polynucleotides
XX CC (AAL07544-AAL26789) and methods of assessing whether a patient is
XX CC afflicted with breast cancer by examining the correlation between the
XX CC expression of certain markers and the cancerous state of breast cells.
XX CC The polynucleotides and encoded polypeptides are potential markers for
XX CC detecting, diagnosing, monitoring, characterising treating and
XX CC potentially preventing breast cancer. The polynucleotides and encoded
XX CC polypeptides are also useful for isolating compounds with cytostatic
XX CC activity.
XX SQ Sequence 862 BP; 255 A; 139 C; 170 G; 292 T; 6 other;

Query Match          9.3%; Score 122.8; DB 22; Length 862;
Best Local Similarity 79.7%; Pred. No. 2.8e-22;
Matches 145; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 985 AGATATGTCATCCATACACACCAAGCAGCTCATTTGATATAGATTGGGTAAGAA 1044
DB 364 aggtatactatccatgcacaaaggcagctcatttgatagactggggcaagaa 423
QY 1045 GATGATTCATCTGTTAATAGGCATCTATGATATGTTATGGGTAAGGATA 1104
DB 424 gatgctcccaattgcttaattggcatctatgaatggatggatgggaagtggaaatgatt 483
QY 1105 AAAATGGATCTGATCTCAGTTTGACACAGAAGATTTTACCTGATGATCCAGATAAGAA 1164
DB 484 aaatggatcctgacctcagctctaacaacacaaaggatttatctatttctgagctggttaa 543
QY 1165 CC 1166
DB 544 cc 545

RESULT 9
AAT42758
ID AAT42758 standard; DNA; 153 BP.
XX AC AAT42758;
XX DT 12-MAR-1997 (first entry)
XX DE Chick CHD-W gene fragment.
XX KW Bird; sex determination; chromodomain-Helicase-DNA binding 1;
XX KW CHD-1A; CHD-W; W chromosome; ss.
XX OS Gallus sp.
XX PH Key Location/Qualifiers
XX FT misc_difference 52..81
XX FT /*tag= a
XX FT /note= "bases 52-81 are a repeat of bases 22-51
XX FT and are ignored in the translated amino
XX FT acid sequence given in Fig 3"
XX PN W09639505-A1.
```

```
XX 12-DEC-1996.
XX 05-JUN-1996; 96WO-GB01341.
XX 06-JUN-1995; 95GB-0011439.
XX (ISIS-) ISIS INNOVATION LTD.
XX Griffiths R, Tiwari B;
XX WPI; 1997-043127/04.
XX P-PSDB; AAW08148.
XX Avian chromodomain-helicase-DNA binding genes determine sex in
XX birds - used for sex determ. and to control sex of progeny
XX Claim 8; Fig 3; 76pp; English.
XX Bases 3855-3977 (AAT42756) of the mouse CHD-1 gene show homology
XX to portions of the chicken CHD-1A (A = Avian) gene (AAT42757),
XX chicken CHD-W (W refers to the W chromosome) gene (AAT42758) and
XX and the great tit CHD-W gene (AAT42759). Translated amino acid
XX sequences of this region are provided in AAW08146-49. The CHD-1A
XX (see also AAT42751) and CHD-W (see also AAT42754-55) genes determine
XX sex in birds and can be used to identify the sex of an embryo,
XX foetus etc. and to manipulate the sex of progeny.
XX Sequence 153 BP; 56 A; 36 C; 31 G; 30 T; 0 other;

Query Match 7.6%; Score 100.4; DB 18; Length 153;
Best Local Similarity 99.0%; Pred. No. 1.1e-16;
Matches 101; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1159 AAGAAACCCAGCTAAGCAGTTACAGACTCGTGCAGATTACCTCAATTAATTAATCTGAAT 1218
Db 52 aagaaacccagcgaagcagttacagaccgctgcagattacctcattaaattactgaat 111
|||||
QY 1219 AAGAGACCTTGCAGAAAGGAAGCAGACAGACTTGCTGGTGCA 1260
Db 112 aaagaccttgcaagaagaaggaagcagagaccttgctggtgca 153
|||||

RESULT 10
AAT42757
ID AAT42757 standard; DNA; 153 BP.
XX
AC AAT42757;
XX
DT 12-MAR-1997 (first entry)
XX
DE Chick CHD-1A gene fragment.
XX
KW Bird; sex determination; chromodomain-Helicase-DNA binding 1;
KW CHD-1A; CHD-W; W chromosome; ss.
XX
OS Gallus sp.
XX
FH Key Location/Qualifiers
FT misc_difference 52..81
FT /*tag= a
FT /note= "bases 52-81 are a repeat of bases 22-51
FT and are ignored in the translated amino
FT acid sequence given in Fig 3"
XX
PN WO9639505-A1.
XX
PD 12-DEC-1996.
XX
PF 05-JUN-1996; 96WO-GB01341.
XX
PR 06-JUN-1995; 95GB-0011439.
XX
```

```
XX (ISIS-) ISIS INNOVATION LTD.
XX Griffiths R, Tiwari B;
XX WPI; 1997-043127/04.
XX P-PSDB; AAW08147.
XX Avian chromodomain-helicase-DNA binding genes determine sex in
XX birds - used for sex determ. and to control sex of progeny
XX Claim 8; Fig 3; 76pp; English.
XX Bases 3855-3977 (AAT42756) of the mouse CHD-1 gene show homology
XX to portions of the chicken CHD-1A (A = Avian) gene (AAT42757),
XX chicken CHD-W (W refers to the W chromosome) gene (see also AAT42758)
XX and the great tit CHD-W gene (see also AAT42759). Translated amino
XX acid sequences of this region are provided in AAW08146-49. The
XX CHD-1A (see also AAT42751) and CHD-W (see also AAT42754-55) genes
XX determine sex in birds and can be used to identify the sex of an
XX embryo, foetus etc. and to manipulate the sex of progeny.
XX Sequence 153 BP; 58 A; 40 C; 31 G; 24 T; 0 other;

Query Match 7.0%; Score 92.4; DB 18; Length 153;
Best Local Similarity 94.1%; Pred. No. 1.3e-14;
Matches 96; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1159 AAGAAACCCAGCTAAGCAGTTACAGACTCGTGCAGATTACCTCAATTAATTAATCTGAAT 1218
Db 52 aagaaacccagcgaagcagttacagaccgctgcagattacctcattaaattactgaat 111
|||||
QY 1219 AAGAGACCTTGCAGAAAGGAAGCAGACAGACTTGCTGGTGCA 1260
Db 112 aaagaccttgcaagaagaaggaagcagagaccttgctggtgca 153
|||||

RESULT 11
AAT42759
ID AAT42759 standard; DNA; 153 BP.
XX
AC AAT42759;
XX
DT 12-MAR-1997 (first entry)
XX
DE Great tit CHD-W gene fragment.
XX
KW Bird; sex determination; chromodomain-Helicase-DNA binding 1;
KW CHD-1A; CHD-W; W chromosome; ss.
XX
OS Parus major.
XX
FH Key Location/Qualifiers
FT misc_difference 52..81
FT /*tag= a
FT /note= "bases 52-81 are a repeat of bases 22-51
FT and are ignored in the translated amino
FT acid sequence given in Fig 3"
XX
PN WO9639505-A1.
XX
PD 12-DEC-1996.
XX
PF 05-JUN-1996; 96WO-GB01341.
XX
PR 06-JUN-1995; 95GB-0011439.
XX
PA (ISIS-) ISIS INNOVATION LTD.
XX Griffiths R, Tiwari B;
XX WPI; 1997-043127/04.
XX
```

DR P-PSDB; AAW08149.

XX Avian chromodomain-helicase-DNA binding genes determine sex in

PT birds - used for sex determ. and to control sex of progeny

XX

XX Claim 8; Fig 3; 76pp; English.

PS

CC Bases 3855-3977 (AAT42756) of the mouse CHD-1 gene show homology

CC to portions of the chicken CHD-1A (A = Avian) gene (AAT42757),

CC chicken CHD-W (W refers to the W chromosome) gene (AAT42758) and

CC and the great tit CHD-W gene (AAT42759). Translated amino acid

CC sequences of this region are provided in AAW08146-49. The CHD-1A

CC (see also AAT42751) and CHD-W (see also AAT42754-55) genes determine

CC sex in birds and can be used to identify the sex of an embryo,

CC foetus etc. and to manipulate the sex of progeny.

XX

SQ Sequence 153 BP; 60 A; 33 C; 31 G; 29 T; 0 other;

Query Match 6.7%; Score 87.6; DB 18; Length 153;

Best Local Similarity 91.2%; Pred. No. 2.4e-13;

Matches 93; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1159 AAGAACCCAGGCTAAGCAGTTACAGACTCGTGCGAGATTACCTCATTAATTAATCTGAAT 1218

DB 52 aagaaacacagggcaagcaggtgcagccgtgcagacttaacttaataactgaat 111

QY 1219 AAAGACCTTGCAGAAAGGAGACACAGAGACTTGCCTGGTGCA 1260

DB 112 aagacacctgcaagaaagagtgcaagacttactgtgca 153

RESULT 12

AAA01247

ID AAA01247 standard; CDNA; 249 BP.

XX

AC AAA01247;

XX

DT 19-MAY-2000 (first entry)

XX

DE Human colon cancer cell line polynucleotide sequence SEQ ID NO:1238.

XX

XX Human; colon cancer; tumour; diagnosis; gene expression product;

KW probe; detection; cancerous state; metastasis; identification;

KW breast cancer; oestrogen receptor-positive breast cancer; therapy;

XX oestrogen receptor-negative breast cancer; lung cancer; ss.

XX

OS Homo sapiens.

XX

PN W09958675-A2.

XX

PD 18-NOV-1999.

XX

PF 13-MAY-1999; 99WO-US10602.

XX

PR 14-MAY-1998; 98US-0085426.

PR 15-MAY-1998; 98US-0085537.

PR 15-MAY-1998; 98US-0085696.

PR 21-OCT-1998; 98US-0105234.

PR 27-OCT-1998; 98US-0105877.

XX

XX (CHIR ) CHIRON CORP.

PA (HYSE-) HYSEQ INC.

XX

XX Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;

PI Reinhard C, Giese K, Randazzo F, Kennedy GC, Pot D, Kassam A;

PI Lanson G, Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Labat I;

XX Leshkowitz D, Kita D, Garcia V, Jones LW, Stache-Crain B;

XX

DR WPI; 2000-126369/11.

XX

XX Polynucleotide library used to determine cancerous states of mammalian

PT cells -

XX Claim 1; Page 480; 1097pp; English.

PS

XX AAA00010 to AAA02716 represent polynucleotides isolated from cDNA

CC libraries constructed from human colon cancer cell lines. The present

CC invention also describes a method of detecting differentially expressed

CC genes correlated with a cancerous state of a mammalian cell, comprising

CC detecting at least one differentially expressed gene product in a test

CC sample derived from a cell suspected of being cancerous, where detection

CC of the differentially expressed gene product is correlated with a

CC cancerous state of the cell from which the test sample was derived.

CC The polynucleotide sequences can be used in a method for detecting

CC differentially expressed genes correlated with a cancerous state of a

CC mammalian cell. The polynucleotides can also be used as probes for

CC detecting and mapping related genes. They can be used in diagnosis and

CC prognosis of diseases and disorders (e.g. identification of

CC pre-metastatic or metastatic cancerous states, stages of cancer, or

CC responsiveness of cancer to therapy). This is particularly for breast

CC cancer, oestrogen receptor-positive breast cancer, oestrogen receptor-

CC negative breast cancer, lung cancer, and colon cancer.

XX

SQ Sequence 249 BP; 71 A; 31 C; 64 G; 61 T; 22 other;

Query Match 6.3%; Score 83.4; DB 21; Length 249;

Best Local Similarity 60.6%; Pred. No. 3.7e-12;

Matches 134; Conservative 0; Mismatches 83; Indels 4; Gaps 1;

QY 168 GTTATCAGCAATTTTGAAGTTTGGTGCTCGAGGAACCTTTTAAAGAACCTGAANNNGAAGA 227

DB 17 gctgacagctattttgaaatttgagcagaggatctctcacaagaacaggagggaggga 76

QY 228 AGAGGAGCCTCAGGAGATGGATATAGATCAATCTGAAGAGCCTGAACTCGAGAAA 287

DB 77 atcagaacctcaggaatgatagataagaaatttcggttgctganacganagagaa 136

QY 288 TGAGTCAGGCCCAATTAACTGTAGGAGATGAGTACTTTTCACAGTTCAAGGTAGCTACTT 347

DB 137 tgaagtgctcancnagtc---ncagatgaanttcacagantaaggttgtnaantt 192

QY 348 TTCCAATATGCGATGAGATGACATTTGAATTGGAACCCAGAAC 388

DB 193 tgcagcnatggangatgngtaactntntcaaaancntggnc 233

RESULT 13

AAT42756

ID AAT42756 standard; DNA; 153 BP.

XX

AC AAT42756;

XX

DT 12-MAR-1997 (first entry)

XX

DE Mouse CHD-1 gene (bases 3855-977).

XX

XX Bird; sex determination; chromodomain-Helicase-DNA binding 1;

KW CHD-1; CHD-W; W chromosome; ss.

XX

OS Mus sp.

XX

XX Key Location/Qualifiers

FT misc\_difference 52..81

FT /tag= a

FT /note= "bases 52-81 are a repeat of bases 22-51

FT and are ignored in the translated amino

FT acid sequence given in Fig 3"

XX

PN W09639505-A1.

XX

PD 12-DEC-1996.

XX

XX 05-JUN-1996; 96WO-GB01341.

XX



PR	20-OCT-2000;	2000US-0241808.
PR	20-OCT-2000;	2000US-0241809.
PR	20-OCT-2000;	2000US-0241826.
PR	01-NOV-2000;	2000US-0244617.
PR	08-NOV-2000;	2000US-0246474.
PR	08-NOV-2000;	2000US-0246475.
PR	08-NOV-2000;	2000US-0246476.
PR	08-NOV-2000;	2000US-0246477.
PR	08-NOV-2000;	2000US-0246478.
PR	08-NOV-2000;	2000US-0246523.
PR	08-NOV-2000;	2000US-0246524.
PR	08-NOV-2000;	2000US-0246525.
PR	08-NOV-2000;	2000US-0246526.
PR	08-NOV-2000;	2000US-0246527.
PR	08-NOV-2000;	2000US-0246528.
PR	08-NOV-2000;	2000US-0246532.
PR	08-NOV-2000;	2000US-0246609.
PR	08-NOV-2000;	2000US-0246610.
PR	08-NOV-2000;	2000US-0246611.
PR	08-NOV-2000;	2000US-0246613.
PR	17-NOV-2000;	2000US-0249207.
PR	17-NOV-2000;	2000US-0249208.
PR	17-NOV-2000;	2000US-0249209.
PR	17-NOV-2000;	2000US-0249210.
PR	17-NOV-2000;	2000US-0249211.
PR	17-NOV-2000;	2000US-0249212.
PR	17-NOV-2000;	2000US-0249213.
PR	17-NOV-2000;	2000US-0249214.
PR	17-NOV-2000;	2000US-0249215.
PR	17-NOV-2000;	2000US-0249216.
PR	17-NOV-2000;	2000US-0249217.
PR	17-NOV-2000;	2000US-0249244.
PR	17-NOV-2000;	2000US-0249245.
PR	17-NOV-2000;	2000US-0249264.
PR	17-NOV-2000;	2000US-0249265.
PR	17-NOV-2000;	2000US-0249297.
PR	17-NOV-2000;	2000US-0249299.
PR	01-DEC-2000;	2000US-0249300.
PR	01-DEC-2000;	2000US-0250160.
PR	01-DEC-2000;	2000US-0250391.
PR	05-DEC-2000;	2000US-0251030.
PR	05-DEC-2000;	2000US-0251988.
PR	05-DEC-2000;	2000US-0256719.
PR	06-DEC-2000;	2000US-0251479.
PR	08-DEC-2000;	2000US-0251856.
PR	08-DEC-2000;	2000US-0251868.
PR	08-DEC-2000;	2000US-0251869.
PR	08-DEC-2000;	2000US-0251989.
PR	08-DEC-2000;	2000US-0251990.
PR	11-DEC-2000;	2000US-0254097.
PR	05-JAN-2001;	2001US-0259678.
XX	(HUMA-*) HUMAN GENOME SCI INC.	
XX	Rosen CA, Barash SC, Ruben SM;	
XX	WPI: 2001-502630/55.	
XX	P-PSDB; AAM93109.	
XX	Polynucleotides encoding digestive system antigens, useful for	
PT	diagnosing, treating, preventing and/or prognosing disorders of the	
PT	digestive system, particularly cancer and cancer metastases -	
XX	Claim 1; SEQ ID NO 1198; 986pp; English.	
XX	The present invention provides the protein and coding sequences of a	
XX	number of human digestive system antigens. These can be used in the	
CC	diagnosis, treatment and prevention of digestive system disorders,	
CC	including cancer, Meckel's diverticulum, bacterial or parasitic	
CC	infections, appendicitis, Hirschsprung's disease, chronic colitis or	
CC	ulcerative colitis. The present sequence is a cDNA encoding a digestive	
CC	system antigen of the invention.	

```
PR 22-AUG-2000; 2000US-0225868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-457727/49.
DR P-PSDB; AAM38625.
XX
PT Isolated polypeptide for treating, preventing and/or prognosing
PT disorders related to the colon and rectum including colorectal cancers
PT and also for testing and detection e.g. diagnosis -
XX
PS Claim 1; SEQ ID NO: 67; 522pp + Sequence Listing; English.
XX
CC The present invention provides the protein and coding sequences of a
CC number of colorectal cancer antigens. These are shown in
CC AA157547-AA157619 and AAM38569-AAM38641. These can be used in the
CC diagnosis, prevention and treatment of cancer of the colon and/or rectum.
CC The present sequence is a colorectal cancer antigen coding sequence of
CC the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 421 BP; 148 A; 78 C; 127 G; 61 T; 7 other;
Query Match 5.0%; Score 65.2; DB 22; Length 421;
Best Local Similarity 60.9%; Pred. No. 2.7e-07;
Matches 106; Conservative 0; Mismatches 68; Indels 0; Gaps 0;
QY 1128 GACACAGAGATTTTACCTGATCCAGATCAAGAAACCCAGGCTTACGAGTACAGAC 1187
Db 100 gactgacaaaattctgccggtgagacagataaaagcctcagggaagcagctacagac 159
QY 1188 TCGTCAGATTACCTCATTAAATTACTGAATAAAGACCTTGCAAGAAAGAGACGACAG 1247
Db 160 ccgagcgggattacttggtaagctgctcagaagggtctggagagaaggggctgtgac 219
QY 1248 ACTTCTGTGCGAGCAATTCAAAGAGGAGAAACCAAGCAAGTAAAGCAAGTAA 1301
Db 220 aggtgggggaaggccaaattaaagaagcggagcctcgggttaagaagagaaaaa 273
```

Mon Aug 5 11:51:46 2002

us-08-973-363-15.rng

Page 13

Search completed: August 2, 2002, 20:53:21  
Job time: 26292 sec

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.  
OM nucleic - nucleic search, using sw model  
Run on: August 2, 2002, 16:53:19 ; Search time 8940.62 Seconds  
(without alignments)  
1986.662 Million cell updates/sec  
Title: US-08-973-363-15  
Perfect score: 1316  
Sequence: 1 ATTATCGGCTAGTCACAAA.....AATAAGCAACAAAGGCTGC 1316  
Scoring table: Gapop 10.0 , Gapext 1.0  
Searched: 13736207 seqs, 6748477542 residues  
Total number of hits satisfying chosen parameters: 27472414  
Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries  
Database : EST: \*  
1: em\_estba: \*  
2: em\_esthum: \*  
3: em\_estin: \*  
4: em\_estnu: \*  
5: em\_estov: \*  
6: em\_estpl: \*  
7: em\_estro: \*  
8: em\_hic: \*  
9: gb\_estl: \*  
10: gb\_est2: \*  
11: gb\_hic: \*  
12: gb\_gss: \*  
13: em\_gss\_hum: \*  
14: em\_gss\_inv: \*  
15: em\_gss\_pln: \*  
16: em\_gss\_vrt: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES				Description	
Result No.	Score	Query Match %	Length	ID	
1	473.2	36.0	1811	AK015218	AK015218 Mus muscu
c 2	459.2	34.9	686	AW997058	AW997058 QV3-BN004
3	437.6	33.3	866	AU125712	AU125712 AU125712
4	415.8	31.6	765	BG485663	BG485663 602505962
5	410	31.2	633	BF427518	BF427518 df89c11.y
6	393.2	29.9	645	AL644594	AL644594 AL644594
7	387.2	29.4	619	BB155356	BB155356 BB155356
8	387.2	29.4	622	AL654479	AL654479 AL654479
9	358	27.2	660	BB461065	BB461065 BB461065
c 10	342.2	26.0	482	AA414473	AA414473 VCS9D12.S
c 11	332.2	25.2	609	AW467566	AW467566 he21e07.y
c 12	314.4	23.3	459	BF816895	BF816895 M63-C1018
c 13	293.8	22.3	593	AL659353	AL659353 AL659353
c 14	268.6	20.4	430	AL601246	AL601246 DKZp313J
15	254.8	20.1	446	BB834922	BB834922 BB834922
16	237.2	18.0	842	AL658477	AL658477 AL658477
c 17	231.2	17.6	337	AW996787	AW996787 QV3-BN004

18	229.2	17.4	643	10	BF611466	BF611466 dd81d08.y
c 19	226.2	17.2	547	9	AI890775	AI890775 vm95f11.x
c 20	220.6	16.8	615	10	BF615108	BF615108 dd81d08.x
21	219.8	16.7	1053	10	BM458124	BM458124 AGNCOURT
22	217.4	16.5	438	9	BB830730	BB830730 BB830730
23	215.2	16.4	1028	10	BE895133	BE895133 601436060
c 24	207.4	15.8	657	10	BM247002	BM247002 K0745H01-
c 25	190.4	14.5	480	10	BF388692	BF388692 UI-R-B52-
c 26	189.8	14.4	521	10	BF396897	BF396897 UI-R-B52-
c 27	188.8	14.3	335	9	AA179333	AA179333 zp47c07.s
28	179	13.6	766	10	BE468732	BE468732 Iphdk0038
29	176.6	13.4	771	10	BI659862	BI659862 603302430
30	173	13.1	434	9	AW068785	AW068785 cn24h09.x
c 31	170	12.9	608	10	BM185005	BM185005 fv16c08.x
c 32	168.2	12.8	364	9	BE082266	BE082266 QV2-BT063
c 33	164.2	12.5	337	9	AI695325	AI695325 wa23c02.x
c 34	161.4	12.3	608	10	BM024825	BM024825 fu73e09.x
c 35	160.2	12.2	574	10	BM185888	BM185888 fu74a09.x
c 36	158.6	12.1	578	10	BM070535	BM070535 fu99h05.x
c 37	156.4	11.9	586	9	BE199465	BE199465 us72c05.x
c 38	156	11.9	534	10	BM220519	BM220519 C0941H07-
39	156	11.9	546	10	BG757172	BG757172 602710541
c 40	154.6	11.7	735	9	AI679813	AI679813 tu66a11.x
* 41	154.2	11.7	655	10	BG018373	BG018373 daa23c11.
c 42	152.2	11.6	361	9	AA823080	AA823080 vw40d10.f
c 43	151.4	11.5	710	9	AI679237	AI679237 tu62a11.x
c 44	143.6	10.9	652	9	AI688976	AI688976 tx9te11.x
c 45	141	10.7	334	9	AI466097	AI466097 vw40d10.y

ALIGNMENTS

AK015218 1811 bp mRNA linear HTC 19-JAN-2002  
Mus musculus adult male testis cDNA, RIKEN full-length enriched  
library, clone:4930428D05:chromodomain helicase DNA binding protein  
1, full insert sequence.

AK015218  
AK015218.1 GI:12853470  
HPC: CAP trapper  
Mus musculus (strain:C57BL/6J) adult male testis cDNA to mRNA,  
clone:lib-RIKEN full-length enriched mouse cDNA library  
clone:4930428D05.

ORGANISM  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE  
1 (sites)  
Carninci,P., and Hayashizaki,Y.  
High-efficiency full-length cDNA cloning  
Meth. Enzymol. 303, 19-44 (1999)  
99279253  
PUBMED 10349636

REFERENCE  
2 (sites)  
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,  
Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new genes  
Genome Res. 10 (10), 1617-1630 (2000)  
20499374  
PUBMED 11043159

REFERENCE  
3 (sites)  
Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,  
Konno,H., Akiyama,J., Nishi,K., Kitsuai,T., Tashiro,H., Itoh,M.,  
Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,  
Yamanoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,  
Fujisake,S., Inoue,K., Togawa,I., Izawa,M., Ohara,E., Watanabe,M.,  
Itoneda,I., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J.,  
Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer  
Genome Res. 10 (11), 1757-1771 (2000)





Db 1 ATAAGAAATTTGGTGTCTCTGTGAAAGATTAGATCAATTTGCTGAGATGCTGAGTTAG 60  
QY 749 TTGATAAATCTGAAACAGACCTTTAGACGCTGTGGGAACTTGTACATAATGATGATGATTA 808  
Db 61 TTGATAGTCAGAAACAGACCTTTAGACGCTGTGGGAACTTGTACATAATGATGATGATTA 120  
QY 809 AGGCTTTAAATGATAATGACTTTGGTCAAGGAGAAACAGTGGTAGATTTGGGAAAGTTA 868  
Db 121 AAGCAATTAAGGATAGTTCTTCAGGAACAGAAACAGTGGTAGACTCGGAAAGTGA 180  
QY 869 AAGGCCAACATCCGAATAGCAGAGTGCAGTGAATGCAAGCTAGTCAATTTCTCAGC 928  
Db 181 AGGGTCCAAATCTCCGAATAGCAGAGTGCAGTGAATGCAAGCTAGTCAATTTCTCAGC 240  
QY 929 AAGAAGAGTTGGCACCATTGCATAAATCGATTCCTTCAGATCCAGAAAGGAAAGAT 988  
Db 241 AAGAAGAAATTAATACCTTTGCACAAATCCATTCCTTCATCCAGAAAGAAAGCAGT 300  
QY 989 ATGTCATCCCATACCAACCAAGCAGCTCATTTTGCATATAGATTTGGGTAAAGAAAGATG 1048  
Db 301 ATACTATCCATCCACACAAAGCAGCTCATTTTGCATATAGATTTGGGTAAAGAAAGATG 360  
QY 1049 ATTCCAAATCTGTTAATAGCATCTATGAATATGTTATGCGAGTTCGGGAAATGATAAAA 1108  
Db 361 ATTCCAAATTTGTTAATGTCATCTATGAATATGTTATGCGAGTTCGGGAAATGATAAAA 420  
QY 1109 TGGATCCTGATCTCAGCTTTGACACAGAAGATTTTACCTGATGATCCAGATAAGAACCC 1168  
Db 421 TGGATCCTGATCTCAGCTTTGACACAGAAGATTTTACCTGATGATCCAGATAAGAACCC 1168  
QY 1169 AGGCTAAGCATTTACAGACTTCGTCAGATTTACCTCATTAATTAATTAAGAAAGCTTTG 1228  
Db 481 AAGCAAAACAGTTGACAGACCCGTCAGACTACCTCATCAATTAATTAAGAAAGCTTTG 540  
QY 1229 CAAGAAAGGAGCAGACAGACTTTGTCAGGCAATTTCAAGAGAGAGAAACCAAGAA 1288  
Db 541 CAAAAAAGAGC-----TCCTTCGTCGCGGAAAGTTCAAAGAGAGAGAAAGCAAGAG 594  
QY 1289 GTAAGAGAAATAAGCAACAAAGGCT 1314  
Db 595 CTAAAGAGAAATAAGCAATGAAGTCT 620

RESULT 4  
BG485663 765 bp mRNA linear EST 21-MAR-2001  
LOCUS 602505962F1 NIH\_MGC\_77 Homo sapiens cDNA clone IMAGE:4619294 5',  
DEFINITION mRNA sequence.  
ACCESSION BG485663  
VERSION BG485663.1 GI:13417942  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 765)  
AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>  
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)  
COMMENT Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue procurement: CLONTECH Laboratories, Inc.  
cDNA Library Preparation: CLONTECH Laboratories, Inc.  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLM1378 row: 1 column: 15  
High quality sequence stop: 593.  
Location/Qualifiers  
1. .765  
/organism="Homo sapiens"

FEATURES  
source

/db\_xref="taxon:9606"  
/clone\_image="4619294"  
/clone\_lib="NIH\_MGC\_77"  
/lab\_host="DH10B (TI phage-resistant)"  
/note="Organ: lung; Vector: pDNR-LIB (Clontech); Site\_1:  
sfll (ggcgctcgcc); Site\_2: sfll (ggcattatggcc); 5' and  
3' adaptors were used in cloning as follows: 5' adaptor  
sequence: 5'-CAGCGCATTTATGGC-3' and 3' adaptor sequence:  
5'-ATTCTAGAGCGGAGCGCGCACATG-DT(30)BN-3' (where B = A,  
C, or G and N = A, C, G, or T). Average insert size 1.9  
kb (range 0.5-4.0 kb). 12/15 colonies contained inserts  
by PCR. This library was enriched for full-length clones  
and was constructed by Clontech Laboratories (Palo Alto,  
CA). Note: this is a NIH\_MGC Library."  
BASE COUNT 272 a 134 c 181 g 178 t  
ORIGIN  
Query Match 31.6%; Score 415.8; DB 10; Length 765;  
Best Local Similarity 82.3%; Pred. No. 1.5e-71;  
Matches 488; Conservative 0; Mismatches 104; Indels 1; Gaps 1;  
QY 272 CTGAAACTCGAGAAATAGTCTAGGCCCATTAATCTAGGAGATGAGTTACTTTTCACAGT 331  
Db 2 CTGAAACTCATGAAATGAACAGAGTCTTAACTCTAGGAGATGAATTCCTTTCCAGT 61  
QY 332 TCAAGTAGCTACTTTTCCATATGATCAAGATGACATTTGATTTGGACCAACACAAA 391  
Db 62 TCAAGTTGCCAACTTCTCAAAATATGATGAGATGACATTTGATTTGGAAACCTTGAAGAA 121  
QY 392 ATCTAAGAACTGGGAGAAATCATTTCCAGAGTTTCAGTGGCGCAGCAATTAAGGGGNGG 451  
Db 122 ATCTAAGAAATCGGAGAGAAATTTCCAGAGATCAAGAAACACCATTTAGAGAGAAG 181  
QY 452 AAAGACAAAGAACTTTGAAGAAATATATATGCTTCCAGAAATGAGAAACTGTGCAAAAC 511  
Db 182 AAAGACAAAGAACTTTGAAGAAATTTATATGCTTCCCAAGAAATGAGAAATTTGTGCAAAAC 241  
QY 512 AGATCAGCTTTAATGGAATGAAGGAGATGTCAGTGGAGGAGGAGGAGGAGGAGGAGTCTG 571  
Db 242 AGATTAGTTCAATGGAAGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGTCTG 301  
QY 572 ATAGTGATTCATCTCAGAAAGAAACGACCAAAAGAAACGTCGACGACCAACCAATATTC 631  
Db 302 ATAGTGATTCATCTCAGAAAGAAAGGCAAGAAACGTCGACGACCAACCAATATTC 361  
QY 632 CCGGTGAAACATTAAGAGTTTATGATGTCAGAGATTTAGACGATTTATCAAGAGTTACA 691  
Db 362 CTCGGGAGAAATATTAAGAGTTTATGATGTCAGAGAAATTTAGCGGTTTATCAAGAGCTATA 421  
QY 692 A-GAAATTTGGTGGCCAGTTGAAGGTTAGATGCTATAGCTAGAGATGCTGAGTAGTT 750  
Db 422 AGAAATTTGGTGGTCTCTGGAAGATTAGATGCAATTTGCTCGAGATGCTGAGTAGTT 481  
QY 751 GATAAATCTGAAACAGACCTTAGACGCTGGGAGAACTTGTACATAATGATGCATTAAG 810  
Db 482 GATACCTCAGAAACACACCTTAGACGCTCGGAGAACTGCTAGATTAATGTTGCATTA 541  
QY 811 GCTTTAAATGATAATGACTTTTGGTCAAGGAGAAACAGGTTGGTATTTGGGAA 863  
Db 542 GCATTAACGATAGTTCTTCAGGAACAGAAACAGGTTGGTAAAGACTCGGAA 594  
RESULT 5  
BF427518  
LOCUS  
DEFINITION  
df89c11.y1 Xenopus laevis oocyte non normalized Xenopus laevis cDNA  
clone IMAGE:3746325 5' similar to TR:042142 042142  
CHROMO-HELICASE-DNA-BINDING ON THE Z CHROMOSOME PROTEIN. ; mRNA  
sequence.  
ACCESSION BF427518  
VERSION BF427518.1 GI:11439058  
KEYWORDS EST.





```
QY 649 GGATTTAGTGATGAGAGATTAGACGATTTATCAAGAGATTACAGAAATTTGGTGCCCA 708
Db 1 GGATTTAGTGATG -GGAATTCGGCGTTTATCAAGAGCTATAGAAATATGTTGGCCCC 59
QY 709 GTTGAAGGTTAGATGCTATAGCTAGAGATGCTAGCTAGTTGATTAATCTGAACAGAC 768
Db 60 CTGAAGGTTAGATGCAATGCTCGATGCTGAATGTTGTTGATGCTAGAGACAGAT 119
QY 769 CTAGAGCTGTGGGAACCTGTACATTAATGGATGATTAAGGCTTTAAATGATTAATGAC 828
Db 120 CTGAAGAGCTAGAGAACTGTGTCATTAATGGTTGTTAAAGCTTTAAAGAGACAGTTCT 179
QY 829 TTTGGTCAAGGAAGACAGGTGTAGATTTGGGAAAGTTAAAGGCCCAACATTTCCCAATA 888
Db 180 TCAGGAACAGAGCGAGCGAGCTGGCAGACTTGGAAAAGTGAAGGGGCCAATTTCCGCATC 239
QY 889 GCAGAGTGCAGGTGAATGCAAGCTAGTCAATTTCTCAGGAAGAGAGATTGGCACCATTG 948
Db 240 TCTGGAGTCCAAGTGAATGCCAAGCTGTGTCATTTGCCCATGAGGATGAGCTGATCCCTCTG 299
QY 949 CATAAATCGATTCCTTCAGATCCAGAAGAAAGGAAAGATATGTCTATCCCATACACACC 1008
Db 300 CATAAGTCCATCCCTTCGGACCCGGAGGAGGAGCAGTATACAAATCCCTCGCATACA 359
QY 1009 AAGCAGCTCATTTTGTATAGATTGGGTAAACAGAGATGATTCATCTGTTAATAGGC 1068
Db 360 AAGCTGCGCACTTTGATATAGCTTGGGCAAGAGATGATTCATTTGTTAATGTT 419
QY 1069 ATCTATGAATGTTATGCGAGTTGGGAATGATTAATGATGATTCAGTTGTTG 1128
Db 420 ATCTATGATGATGCTATGGAAGCTGGGAATGATTAATGATGATTCAGTTGTTG 479
QY 1129 ACACAGAAGATTTACCTGATGATCCAGATTAAGAAACCCAGGCTTAAGCAGTTACAGCT 1188
Db 480 ACACAGAAGATTTCTCCAGATGATCTGTATAAACCAACCAAGCAACAGATTAACAGACC 539
QY 1189 CGTGCAGATTACCTCATTAATTAATGATTAAGACCTTGCAGAAAGAGGAGCAGACAGAG 1248
Db 540 CGTGCAGATTACCTCATCAAACTTACTTAGCAGAGATCTTGCAGAAAGAGGAGGCTCAGAG 599
QY 1249 CTTGCTGGTGAGGCAATTC 1268
Db 600 CTTTGTGGTGCGGAGGTTTC 619
```

```
RESULT 8
AL654479 AL654479 622 bp mRNA linear EST 13-DEC-2001
LOCUS AL654479 XGC-gastrula silurana tropicalis cDNA clone Tgas044c06 5',
DEFINITION mRNA sequence.
ACCESSION AL654479.1 GI:17665937
VERSION AL654479.1
KEYWORDS EST.
SOURCE western clawed frog.
ORGANISM Silurana tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
Xenopodinae; Silurana.
1 (bases 1 to 622)
Huckle,E., Taylor,R., Ashurst,J.L., Zorn,A.M. and Rogers,J.
AUTHORS Huckle Xenopus tropicalis EST project 2001 (10_2001).
TITLE Unpublished (2001)
JOURNAL Contact: Huckle E
COMMENT Sanger Centre
Hinnton, Cambridgeshire, CB10 1SA, UK
Email: trop@sanger.ac.uk
Sanger Xenopus tropicalis EST project 2001
TROPICALIS_SEQUENCE_ID: Tgas044c06.sp6
Sequencing primer: SP6
This sequence is from a Xenopus Gene Collection (XGC) library
constructed by Aaron M. Zorn.
Location/Qualifiers
source 1..622
```

```
/organism="Silurana tropicalis"
/db_xref="taxon:8364"
/clone="Tgas044c06"
/clone_lib="XGC-gastrula"
/dev_stage="gastrula (stages 10.5-13 mixed)"
/lab_host="Escherichia coli XL1-blue"
/notes="vector: pCS107; Site_1: EcoRI; Site_2: NotI; cDNA
was oligo dr primed from Sug of poly A+ RNA from stages
10-13 gastrulae. EcoRI-NotI cut cDNA was then ligated
into pCS107 with EcoRI at the 5' end and NotI at the 3'
end."
```

BASE COUNT 210 a 115 c 155 g 142 t

ORIGIN

Query Match 29.4%; Score 387.2; DB 9; Length 622;  
Best Local Similarity 80.0%; Pred. No. 5.9e-66;  
Matches 493; Conservative 0; Mismatches 118; Indels 5; Gaps 3;

```
QY 517 AGCTTTAATGGAATGAAGGAGATGCAGTAGGAGCAGAAAGATATTCTGGATCTGATAGT 576
Db 12 AACTTTAATGGAAGTGAAGGTAGACGCAGCAGAACAGAG-TACTCTGGATCTGACAGT 70
QY 577 GATTCCATCTCAGAAAGAACACCAAAACGTCGACGACCAAGCTATTTCGCCGT 636
Db 71 GATTCCATCTCAGTAAGAAACCCCAAAAGAG-AGTCGGCCAGCAACATTTCCAAGA 129
QY 637 GAAACATTAAGGATTTAGTGATGAGAGATGAGAGATTTATCAAGAGTTACAAGAA 696
Db 130 GAAATATTAAGAGGTTTGTAGTGTGATGCTGAGATGAGAGGTTTATTAAGGCTACAAAAG 189
QY 697 TTTGGTGCCCGCTTGAAGGTTAGATGCTATAGCTAGAGATGCTGAGCTAGTTGATPAA 756
Db 190 TTTGGTGCCCGCTTGAAGGTTAGATGCTATAGCTAGAGATGCTGAGTTGGTGGACAAG 249
QY 757 TCTCAACAGACCTTTAGACGCTCTGGGAGAACCTTGTAGATTAATGGATGATTAAGGCTTTA 816
Db 250 TCTGAGACTGATCTCGCGAGTTGAGAGATTAATACACAGCTGCTGCTGATTAAGGCTTTG 309
QY 817 AATCATTAATGACTTTGGTCAAGGAAGACAGAGTGTAGATTTGGGAAAGTTTAAAGGCCCA 876
Db 310 AAGGATAGT---TCTGGGCAAGAAAGGCGAGGAGAGAGTGGAAATGAAAGGACCA 366
QY 877 ACATTCGATAGAGAGTGCAGGTCGAATGCAAGCTAGCTATTTCTCAGCAAGAGAGAG 936
Db 367 ACATTCGATAGTGCAGGTCGAATGCAAGCTAGCTATTTCTCAGCAAGAGAGAG 426
QY 937 TTGGCAGCATTTGATTAATGATTCCTTCAGATCCAGAAAGAAAGATATGTCATC 996
Db 427 CTAGCTCTTTACAAAATCTATTCCTTCAGATCCGATGAAAGAAAGATATGTTATC 486
QY 997 CCATACACACCAAGAGCAGCTCATTTTGTATATAGATTTGGGGTAAAGAGATGATTCAT 1056
Db 487 CCATGCGCACACCAAGAGCTGCTCATTTTGTATATTTGACTGGGGCAAGAGATGACTCAAC 546
QY 1057 CTGTTAATGAGCATCTATGAATGTTTATGGCAGTTGGGAAATGATAAAATGATTCCT 1116
Db 547 CTCTTAATGGCATCTATGAATATGGTATGCGAGCTGGGAAATGATAAAATGAGCCT 606
QY 1117 GATCTCAGTTTGACAC 1132
Db 607 GACCTCAACCTAACAC 622
```

RESULT 9

```
BB461065 BB461065 660 bp mRNA linear EST 25-OCT-2001
LOCUS BB461065 RIKEN full-length enriched, 12 days embryo spinal ganglion
DEFINITION Mus musculus cDNA clone D130070B13 3', similar to I10410 Mouse
cDNA-binding protein (CHD-1) mRNA, mRNA sequence.
ACCESSION BB461065
VERSION BB461065
KEYWORDS EST.
```

SOURCE  
ORGANISM

house mouse.  
Mus musculus  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
1 (bases 1 to 660)

REFERENCE  
AUTHORS

Arakawa, T., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A.,  
Hiramoto, K., Ishii, Y., Ito, M., Kawai, J., Konno, H., Kouda,  
M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, K., Ohno, M.,  
Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki,  
D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H.,  
Tagami, M., Tagawa, A., Takahashi, F., Takeda, Y., Tanaka, T., Toya, T.,  
Muramatsu, M. and Hayashizaki, Y.

TITLE  
JOURNAL

RIKEN Mouse ESTs (Arakawa, T., et al. 2001)

## COMMENT

Unpublished (2001)

On Jul 21, 2000 this sequence version replaced gi:9356558.

Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216

Email: genome-res@gsr.riken.go.jp,  
URL: <http://genome.gsc.riken.go.jp/>  
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh,  
M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. Genome Res. 10 (10), 1617-1630 (2000)  
wagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,  
Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura,  
S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and  
Hayashizaki, Y.

RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
10 (11), 1757-1771 (2000)  
Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara,  
Y. and Hayashizaki, Y.

Computer-based methods for the mouse full-length cDNA  
encyclopedia: real-time sequence clustering for construction of a  
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
Kondo, S., Shinagawa, A., Saito, T., Kiyosawa, H., Yamana, I., Aizawa,  
K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K. and  
Hayashizaki, Y.

Computational Analysis of Full-Length Mouse cDNAs Compared with  
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)  
Please visit our web site (<http://genome.gsc.riken.go.jp/>) for  
further details.

cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.

FEATURES  
source

Location/Qualifiers  
1..660  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="D130070B13"  
/clone\_lib="RIKEN full-length enriched, 12 days embryo  
spinal ganglion"  
/tissue\_type="spinal ganglion"  
/dev\_stage="12 days embryo"  
/lab\_host="DH10B"  
/note="Site\_1: SalI; Site\_2: BamHI; cDNA library was  
prepared and sequenced in Mouse Genome Encyclopedia  
Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in  
RIKEN. Division of Experimental Animal Research in Riken  
contributed to prepare mouse tissues. 1st strand cDNA was  
primed with a primer [5'  
GAGAGAGAGCGGCCGCACTCGAGTTTCTTTTCTTTT 3'], cDNA was  
prepared by using trihalose thermo-activated reverse  
transcriptase and subsequently enriched for full-length by

cap-trapper. Second strand cDNA was prepared with the  
primer adapter of sequence [5'  
GAGAGAGAGATTCTCGAGTTAATTAATATCCCCCCCCCC 3']. cDNA  
was cleaved with BamHI and XhoI. Vector: a modified  
pBluescript KS(+) after bulk excision from Lambda FLC I."

BASE COUNT 211 a 130 c 168 g 151 t  
ORIGIN

Query Match 27.2%; Score 358; DB 9; Length 660;  
Best Local Similarity 74.4%; Pred. No. 3e-60;  
Matches 488; Conservative 0; Mismatches 145; Indels 23; Gaps 2;

680 TCAAGAGTTTACAAGAAATTTGCTGGCCAGTTGAAAGGTTAGATGCTATGCTAGAGATG 739  
Db 2 TCAAGAGCTAATAGTATGTTGGTGGCC-CCTGGAGGTTAGATGCAATGCTCGAGATG 60  
Qy 740 CTGAGCTAGTTGATAAATCTGAAACAGACCTTTAGACGCTTGGGAGAACTTTACATAATG 799  
Db 61 CTGAATTTGGTTGATAGTCAGAAACAGATCTCAGAAAGACTAGGAGAACTGGTGCAATG 120  
Qy 800 GATGCATTAAGCTTTAATGATATGACTTTTGGTCAAGGAAGACAGAGTGGTAGATTG 859  
Db 121 GTTGTGTTAAAGCTTTAAAGAGCAGTCTTTCAGGAACAGAGCAGCAGTGGCAGACTTG 180  
Qy 860 GGAAAGTTTAAAGGCCCAACATTCGGAATAGCAGAGTGCGAGTGAATGCAAAAGCTAGTCA 919  
Db 181 GAAAGAGTGAAGGGGCCAATCTCCGATCTCTGGAGTCCAAAGTGAATGCCAGCTGGTCA 240  
Qy 920 TTTCTCAGCAAGAAGAGTTGGCACCATTGCAATAATTCGATTCTTTCAGATCCAGAGAAA 979  
Db 241 TTGCCCATGAGGATGAGCTGATCCCTCTGCAATAGTCCATCCCTTCGGACCCGGAGGAGA 300  
Qy 980 GGAAAGATATGTCATCCCATACACACACAGAGAGCTCATTTGTATATAGATGGGTA 1039  
Db 301 GGAAGCAGTATACATCCCTGCCATACAAAGGCTGCGCAGCTTTGATATAGACTGGGGCA 360  
Qy 1040 AAGAAGATGATCCCAATCTGTTAATAGGCATCTATGAATATGTTATGGCAGTTGGGAAA 1099  
Db 361 AAGAAGATGATCTAATTTGTTAATTTGTTATCTATGATGCTATGCAAGCTGGGAAA 420  
Qy 1100 TGATAAATATGATCCTGATCTAGTCTAGTTTGACACAGA-----AG 1137  
Db 421 TGATTTAAATATGATGATCAGACCTCAGTTTAAACACACAAAGCCTCCATGCTTCTTTTCAG 480  
Qy 1138 ATTTTACCTGATGATCCAGATAAGAAACCCAGGCTAAGCAGTTACAGACTCGTCAGAT 1197  
Db 481 ATTTCTCAGATGATCTGATATAAAACCAAGCAAAACAGTTACAGCCCGTCAGAG 540  
Qy 1198 TACCTCATTAATTTACTGAATAAAGACCTTGCAGAAAGAGAGACAGACAGACTTGTGCT 1257  
Db 541 TACCTCATCAAACTACTTAGCAGATCTTGCAGAAAGAGAGGCTCAGAGACTTTGTGCT 600  
Qy 1258 GCAGCAATTTCAAGAGGAGAGAAAACAGAGTAAAGAGAAATAAGCAACAAGGC 1313  
Db 601 GCGGGAGGTTCAAGAGGAGAGAAAACGAGAGCTAAAGAGAGTAAGCAATGAAGTC 656

RESULT 10  
AA414473  
LOCUS  
DEFINITION  
IMAGE:778871.5' similar to gb:X66028 m.musculus mRNA KYBP (MOUSE);,  
mRNA sequence.  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
AA414473  
vc59d12.s1 Knowles Solter mouse 2 cell Mus musculus cDNA clone  
GI:2074589  
house mouse.  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 482)  
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,



TITLE	JOURNAL	COMMENT
1. The Role of the State in the Development of the Economy	Journal of Economic Surveys	1990, Vol. 4, No. 1, pp. 1-20
2. The Impact of the State on the Development of the Economy	Journal of Economic Surveys	1990, Vol. 4, No. 1, pp. 21-30
3. The Role of the State in the Development of the Economy	Journal of Economic Surveys	1990, Vol. 4, No. 1, pp. 31-40
4. The Impact of the State on the Development of the Economy	Journal of Economic Surveys	1990, Vol. 4, No. 1, pp. 41-50
5. The Role of the State in the Development of the Economy	Journal of Economic Surveys	1990, Vol. 4, No. 1, pp. 51-60
6. The Impact of the State on the Development of the Economy	Journal of Economic Surveys	1990, Vol. 4, No. 1, pp. 61-70
7. The Role of the State in the Development of the Economy	Journal of Economic Surveys	1990, Vol. 4, No. 1, pp. 71-80
8. The Impact of the State on the Development of the Economy	Journal of Economic Surveys	1990, Vol. 4, No. 1, pp. 81-90
9. The Role of the State in the Development of the Economy	Journal of Economic Surveys	1990, Vol. 4, No. 1, pp. 91-100
10. The Impact of the State on the Development of the Economy	Journal of Economic Surveys	1990, Vol. 4, No. 1, pp. 101-110

BASE COUNT  
ORIGIN

191	GTGCTGAGGAAC	TTTTAAGAACCTGGAANNNGAAGAAGAGGAGCCTCAGAGAGATGGATA	250	
200	GTGCTGAGGAGAGCT	TTTTAAGAGCCTTGAGGAGAGGACGAGGAGCCACAGAGAAATGGATA	60	
251	TAGATGAATCCTG	AAGAGGNC	TGAATCGAGAAATGAGTCAGGCCCAATTAAGCTGAG	310
61	TCGATGAATCCTG	AAGAGAGCCGAGACTCAGAGAACGAGCAGGCCCACTGAGTGTGG	120	
311	GAGATGAGTTACT	TTTCCACAGTTCAAGGTAGCTTAAC	TTTTTCCAAATATGATGAAGAGATGACA	370
121	GGGATGAGCTGCT	TTCTCAGTTTAAAGTTTGGCAATTTTCCAAATATGATGAAGATGACA	180	
371	TTGAAATTGGAAC	CACGAACAAATCTAGAAACTGGGAGAAATCAATCCAGAGATTCAGT	430	
181	TTGAAATTGGAAC	CTGGAAGAAATCAAGAGAACTGGGAGGAGATCAATCCAGAGACCAA	240	
431	GGCGAGTAATGAG	GGGNGGAAGACAAAGAACTTGAGAGAAATATATATGCTTCAA	490	
241	GAGCGGCACTAG	AGAGAGGAGACAAAGAACTCGAGAGAAATTTATATGCTCCGGA	300	
491	GAATGAGAACTGT	CCAAACAGATCAGCTTTAATGGAAATGAAGGGAGATGCAGTAGGA	550	
301	GAATGAGAACTGT	CCAAACAGATAAGTTTCATGGAGTGAAGGGAGGGGAGTAGAA	360	
551	GCAGAGATATTC	TGGATCTGATAGTGATTCATCTCAGAAAGAAACGACCAAAAAAC	610	
361	GCAGAGATATTC	TGGATCTGATAGTTCATCTCGAAAGAAACGCGCCGAAGAAC	420	
611	GTGGACGACCAG	CAACTATTTCCCGCTGAAAAACATTTAAAGGATTTAGTGCAGAGATTA	670	

FEATURES	SOURCE
----------	--------

BASE COUNT

QY	1 A
D <sub>b</sub>	445 A
QY	61 A
D <sub>b</sub>	385 A
QY	121 C
D <sub>b</sub>	325 C
QY	181 T
D <sub>b</sub>	265 T

```
QY 241 GAGATGGATATAGTGAATCCTGAAGAGGNCCTGAAACTCGAGAAATGAGTCAGGCCCA 300
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 205 GAAATGGATATAGTGAATCCTGAAGAGGNCCTGAAACTCGAGAAATGAGTCAGGCCCA 146
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 301 TTAAGTGTAGGAGAGTGAATCCTGACAGTTCAGGTCAGGTCAGGTCAGGTCAGGTCAG 360
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 145 TTAAGTGTAGGAGAGTGAATCCTGACAGTTCAGGTCAGGTCAGGTCAGGTCAGGTCAG 86
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 361 GAAGATGACATTCGAACTGACCAACCAAAATCTAAGAACTGGGAGAAATCATTCACA 420
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 85 GAGGATGACATTCGAACTGACCAACCAAAATCTAAGAACTGGGAGAAATCATTCACA 26
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 421 GAAGTTCAGTGGCGACGATAGAGG 445
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 25 GAAGATCAAGAGAGCATTAGAG 1
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 12
BF816895/c
LOCUS BF816895 459 bp mRNA linear EST 13-JAN-2001
DEFINITION MR3-CI0184-081200-002-h01 CI0184 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF816895
VERSION BF816895.1 GI:12153105
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 459)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=MR3&t2=MR3-CI0184-
081200-002-h01&t3=2000-12-08&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 458.
FEATURES
Location/Qualifiers
1..459
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="CI0184"
/dev_stage="Adult"
/note="organ: colon_ins; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT 106 a 112 c 82 g 159 t
ORIGIN
```

Query Match 23.9%; Score 314.4; DB 10; Length 459;  
Best Local Similarity 79.9%; Pred. No. 1e-51;

```
Matches 366; Conservative 0; Mismatches 92; Indels 0; Gaps 0;
QY 97 GACACCACAGGAAAACCTGCTACTACATACAGGCTCTACTCTCTCAAGCTCAACACCTTTT 156
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 459 GCATCACTGGGAAGACAGTACTACATACAGGTTCTGCCCCATCAAGTTCTACTCCTTC 400
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 157 AATAAGGAGAGTTCATCAGCAATTTTGAAGTTTGGTGTGCTGAGGAACCTTTTAAAGAACCT 216
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 399 CATAAAGAAGAGTTCATCAGCCATTTTAAAGTTTGGTGTGCTGAGAACTTTTAAAGAACCT 340
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 217 GAANNNGAAGAGAGGAGGCTCAGAGATGGATATAGATGAATCCTGAAGAGGCTGAA 276
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 339 GAAGAGAGAGAACCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 280
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 277 ACTCGAGAAAATGAGTCAGGCCCATTTAACTGTAGGAGATGAGTTACTTTTCACAGTTCAAG 336
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 279 ACTCATGAAGATGAACCCGGTCTTTAACTGTAGGAGATGAATTCCTTCCAGTTTCAGG 220
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 337 GTAGCTAACTTTTCCAAATATGATGAAGATGACATTTGAATTTGAAACCAAGAGAAAATCTA 396
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 219 GCTGCCCACTTCTCCAATATGATGAGGATGACCTTTGAGTTGGAACCTGAAAGAAATTC 160
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 397 AGAAACTGGGAAGAAATCATTCAGAAAGTTTCAAGTCGCGAGCAATAGAGGGGNNGAAAGA 456
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 159 AAGAAATGGGAGGAAATTTATCCAGAAGATCCAAGAAGACGATTAGAAGAAGAAAGA 100
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 457 CAAAAGAACTTTGAAGAAATATATATGCTTCCAAGAATGAGAAACTGTGCAAAACAGATC 516
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 99 CCAAGGAAGACTTTGAAGAAATTTATGCTCCCCAGAAATGAGAAATTTGTCCTCAACAGATT 40
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 517 AGCTTTAATGAAATGAAGGGAGATGCAGTAGGAGCAG 554
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 39 AGTTTCCATGGAAGTGAAGGGAGGCGCCCTAGAGAAGTAG 2
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 13
AL659353/c
LOCUS AL659353 593 bp mRNA linear EST 13-DEC-2001
DEFINITION AL659353 XGC-neurula Silurana tropicalis cDNA clone TNeu045e20 5',
mRNA sequence.
ACCESSION AL659353
VERSION AL659353.1 GI:17672995
KEYWORDS EST.
SOURCE western clawed frog.
ORGANISM Silurana tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
Xenopodinae; Silurana.
REFERENCE 1 (bases 1 to 593)
AUTHORS Huckle,E., Taylor,R., Ashurst,J.L., Zorn,A.M. and Rogers,J.
TITLE Sanger Xenopus tropicalis EST project 2001 (10_2001)
JOURNAL Unpublished (2001)
COMMENT Contact: Huckle E
Sanger Centre
Hinxton, Cambridgeshire, CB10 1SA, UK
Email: trop@sanger.ac.uk
Sanger Xenopus tropicalis EST project 2001
TROPICALIS SEQUENCE ID: TNeu045e20.sp6
Sequencing primer: SP6
This sequence is from a Xenopus Gene Collection (XGC) library
constructed by Aaron M. Zorn.
Location/Qualifiers
1..593
/organism="Silurana tropicalis"
/db_xref="taxon:8364"
/clone="TNeu045e20"
/clone_lib="XGC-neurula"
/dev_stage="neurula"
/lab_host="Escherichia coli DH10B"
/note="Vector: pCS107; Site_1: EcoRI; Site_2: NotI; cDNA
was oligo dT primed from Sug of poly A+ RNA from neurula.
EcoRI-NotI cut cDNA was then ligated into pCS107 with
EcoRI at the 5' end and NotI at the 3' end."
```

FEATURES  
source

BASE COUNT 119 a 134 c 114 g 226 t  
ORIGIN

Query Match 22.3%; Score 293.8; DB 9; Length 593;  
Best Local Similarity 80.7%; Pred. No. 1.1e-47;  
Matches 343; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

QY 891 AGGAGTGCAGTGAATGCAAGCTAGTCTTCTCAGCAAGAGAGTGGCCACCAATGCA 950  
DB 593 AGGGTGCAGTGAATGCAAGCTAGTCTTCTCAGCAAGAGAGTGGCCACCAATGCA 534  
QY 951 TAAATCGATCTCTCAGATCCAGAGAAAGAAAGATATGTCATCCATACCAACCA 1010  
DB 533 CAATCTATCTCTCAGATCCAGATCCAGAGAAAGAAAGATATGTCATCCATACCAACCA 474  
QY 1011 AGCAGCTCATTTGATATAGATTTGGGGTAAAGAGATGATTCCTGTTAATAGGCAT 1070  
DB 473 AGCTCTCATTTGATATGATTTGACTGGGGCAAGAGATGACTCAAACTCTTAATTGGCAT 414  
QY 1071 CTATGAATATGGTTATGGCAGTTGGGAAATGATATAAATGGATCCTGATCTTGGAC 1130  
DB 413 CTATGAATATGGTTATGGCAGTTGGGAAATGATATAAATGGATCCTGATCTTGGAC 354  
QY 1131 ACAGAAGATTTTACCTGATGATCCAGATAAGAAACCCAGGCTAAGCAGTTACAGACTCG 1190  
DB 353 ACAGAAGATTTTACCTGATGATCCAGATAAGAAACCCAGGCTAAGCAGTTACAGACTCG 294  
QY 1191 TGCAGATFACCTATTAAATTAAGTAAAGACCTTGCAGAAAGAGGACGACGACT 1250  
DB 293 AGCTGACTACCTCATTTAACTTCTCAATAAAGATCTGCTTAGGAAGAGCAAGACT 234  
QY 1251 TGTCTGTCAGGCAATTCAGAGAGAGAGAAACCAAGAGTAAAGAGTAAAGCAACAA 1310  
DB 233 TTCTAGTGGAGTAATCTTAAGAGGAGAAACCAAGGCAAGAGTAAAGTAAAGTGA 174  
QY 1311 GGCTG 1315  
DB 173 ACCTG 169

RESULT 14  
AL601246 430 bp mRNA linear EST 14-AUG-2001  
LOCUS DKFZp313J1040\_r1 313 (synonym: hlcc2) Homo sapiens cDNA clone  
DEFINITION DKFZp313J1040 5', mRNA sequence.

ACCESSION AL601246  
VERSION AL601246.1 GI:15164752  
KEYWORDS EST.  
SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 430)  
Bloeker, H., Boecker, M., Brandt, P., Mewes, H.W., Weil, B. and Wiemann, S.

TITLE EST (Bloeker, H., Boecker, M., Brandt, P., Mewes, H.W., Weil, B. and Wiemann, S.)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Bloecker H  
MIPS

Am Klopferspitz 18a D-82152 Martinsried, Germany  
This is the 5' sequence of the clone insert  
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer  
Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de;  
Sequenced by GfE (National Research Centre for Biotechnology Ltd.,  
Braunschweig/Germany) within the cDNA sequencing consortium of the  
German Genome Project.  
No sl sequence available.

This clone (DKFZp313J1040) is available at the RZPD in Berlin.  
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059  
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.  
Location/Qualifiers

FEATURES

source 1. .430  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="DKFZp313J1040"  
/clone\_lib="313 (synonym: hlcc2)"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/note="Vector: pTriplex2; Site\_1: SfiIA; Site\_2: SfiIB;  
cDNA-collection"  
BASE COUNT 163 a 81 c 86 g 100 t  
ORIGIN

Query Match 20.4%; Score 268.6; DB 9; Length 430;  
Best Local Similarity 82.2%; Pred. No. 9.1e-43;  
Matches 323; Conservative 0; Mismatches 64; Indels 6; Gaps 1;

QY 922 TCTCAGCAAGAGAGTTGGCACCATTGCATTAATCGATTCTCTCAGATCCAGAGAAAGG 981  
DB 2 TCCCATGAAGAGAAATTAATACCTTGCACAAATCCATCTCTGTATCCAGAGAAAGA 61  
QY 982 AAAAGATATGTCATCCCATACCAACACCAAGCAGCTCATTTTGATATAGATGGGTAA 1041  
DB 62 AAGCAGTATCTATCCCATGCCACACAAAGCAGCTCATTTTGATATAGATGGGTAA 121  
QY 1042 GAAGATGATTCCTCAATCTGTTAATAGGCATCTATGAATATGTTATGGCAGTTGGAAATG 1101  
DB 122 GAAGATGATTCCTCAATCTGTTAATAGGCATCTATGAATATGTTATGGCAGTTGGAAATG 181  
QY 1102 ATAAATGGATCTGATCTCAGTTTGACACAGAGAGATTTTACCTGATCCAGATAAG 1161  
DB 182 ATAAATGGATCTGATCTCAGTTTGACACAGAGAGATTTTACCTGATCCAGATAAG 241  
QY 1162 AAACCCAGGCTCTGATCTACAGCTCGTCAGATCTCTCTAATTAATTAATTAATTA 1221  
DB 242 AAACCCAGGCTCTGATCTACAGCTCGTCAGATCTCTCTAATTAATTAATTAATTA 301  
QY 1222 GACTTTGCAAGAGAGGAGCAGACAGAGCTGCTGTCAGGCAATTCAGAGAGAGAGAA 1281  
DB 302 GACTTTGCAAGAGAGGAGCAGACAGAGCTGCTGTCAGGCAATTCAGAGAGAGAGAA 355  
QY 1282 ACAGAGATGAAGAGATTAAGCAACAAGGCT 1314  
DB 356 GCAAGAGCTAAGAGAGATTAAGCAACAAGGCT 388

RESULT 15  
BB834922  
LOCUS BB834922  
DEFINITION BB834922.1 GI:17013165

ACCESSION BB834922  
VERSION BB834922  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 446)  
Akinura, T., Arakawa, T., Carninci, P., Furuno, M., Hanagaki, T., Hayashida, N., Hiramoto, K., Hirose, T., Hirose, T., Imoto, K., Ishii, Y., Ito, M., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Matsuyama, T., Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagawa, A., Takahashi, F., Takaku-Akashira, S., Tanaka, T., Tomaru, A., Toya, T., Watahiki, A., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.  
RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura, T., et al. 2001)

Unpublished (2001)  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic





**THIS PAGE BLANK (USPTO)**

Result No.	Score	Query Match	Length	DB	ID	Description
<p>8</p>						

RESULT	1	A58696	Sequence 15 from Patent WO9639505.	1316 bp	DNA	linear	PAT 06-MAR-1998
LOCUS		A58696					
DEFINITION		A58696					
ACCESSION		A58696					
VERSION		A58696.1	GI:3714253				
KEYWORDS			unidentified.				
SOURCE			unidentified				
ORGANISM			unclassified.				
REFERENCE			1 (bases 1 to 1316)				
AUTHORS			Griffiths, R. and Tiwari, B.				
TITLE			AVIAN GHD GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN BIRDS				
JOURNAL			Patent: WO 9639505-A 15 12-DEC-1996;				
COMMENT			ISIS INNOVATION (GB)				
FEATURES			Other publication AU 5906996 961224.				
source			Location/Qualifiers				
			1. .1316				
			/organism="unidentified"				
			/db_xref="taxon:32644"				
BASE COUNT			493 a	205 c	308 g	304 t	6 others
ORIGIN							

Query Match 99.5%; Score 1310; DB 6; Length 1316;  
Best Local Similarity 100.08; Pred. No. 2.5e-284;  
Matches 1316; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	ATTATCGGCTAGTCACAAAAGGATCAGTAGAAGATATTTCTTGAAAGAGCCCAAGAAA	60
DB	1	ATTATCGGCTAGTCACAAAAGGATCAGTAGAAGATATTTCTTGAAAGAGCCCAAGAAA	60
QY	61	AGAGTGGTGTAGATCATTTAGTGATTCAGAGAAATGGACACACAGGAGGAAATCTGACTA	120
DB	61	AGAGTGGTGTAGATCATTTAGTGATTCAGAGAAATGGACACACAGGAGGAAATCTGACTA	120
QY	121	CATACAGGCTCTACTCCTTCAAGCTCAACACCTTTTAAAGAACCTTGAANNNGAAGAGAGGAGGCTCTCAG	240
DB	121	CATACAGGCTCTACTCCTTCAAGCTCAACACCTTTTAAAGAACCTTGAANNNGAAGAGAGGAGGCTCTCAG	240
QY	181	TTGAAGTTTGGTGGTGGAGAACTTTTAAAGAACCTTGAANNNGAAGAGAGGAGGCTCTCAG	360
DB	181	TTGAAGTTTGGTGGTGGAGAACTTTTAAAGAACCTTGAANNNGAAGAGAGGAGGCTCTCAG	360
QY	241	GAGATGGATATAGATGAATCTGAAGAGGNCCTGAAATCGAGAAATGAGTCAGGCCCA	300
DB	241	GAGATGGATATAGATGAATCTGAAGAGGNCCTGAAATCGAGAAATGAGTCAGGCCCA	300
QY	301	TTAACTGTAGGAGATGAGTTACTTTTCCACAGTTCAAGGTAGCTAACTTTTCCAAATATGGAT	360
DB	301	TTAACTGTAGGAGATGAGTTACTTTTCCACAGTTCAAGGTAGCTAACTTTTCCAAATATGGAT	360
QY	361	GAGATGACATTTGAATTTGAACCCAGAACCAAAATCTAAGAAATCTGGAAGAAATCATTTCCA	420
DB	361	GAGATGACATTTGAATTTGAACCCAGAACCAAAATCTAAGAAATCTGGAAGAAATCATTTCCA	420
QY	421	GAGTTTCAGTGGCGGCAATAGAGGGGNGGAAAGACAAAAGAACTTGAAGAAATATATAT	480
DB	421	GAGTTTCAGTGGCGGCAATAGAGGGGNGGAAAGACAAAAGAACTTGAAGAAATATATAT	480
QY	481	ATGCTTCCAAAGATGAGAACTGTGCAAAACAGATCAGCTTTAAATGGAATGAAGGGAGA	540
DB	481	ATGCTTCCAAAGATGAGAACTGTGCAAAACAGATCAGCTTTAAATGGAATGAAGGGAGA	540
QY	541	TGCGATGAGGACAGAGATATTTCTGGATCTGATAGTGTATTCATCTCAGAAAAGAAACGA	600
DB	541	TGCGATGAGGACAGAGATATTTCTGGATCTGATAGTGTATTCATCTCAGAAAAGAAACGA	600
QY	601	CCAAAAACGTTGGACGACCCAGCACTATTCCTCCGTGAAACATTAAGAGATTAGTGAT	660
DB	601	CCAAAAACGTTGGACGACCCAGCACTATTCCTCCGTGAAACATTAAGAGATTAGTGAT	660
QY	661	GCAGAGATTAGACGATTTATCAAGAGTTTACAAGAAATTTGGTGGCCCACTTGAAGGTTA	720
DB	661	GCAGAGATTAGACGATTTATCAAGAGTTTACAAGAAATTTGGTGGCCCACTTGAAGGTTA	720
QY	721	GATGCTATAGCTAGAGATGCTGAGCTAGTGTATTAATCTGAAACAGACCTTGAAGCTCTG	780
DB	721	GATGCTATAGCTAGAGATGCTGAGCTAGTGTATTAATCTGAAACAGACCTTGAAGCTCTG	780
QY	781	GGAGAACTTGTACATAATCGATTAAGCTTTAAATGATTAAGCTTTGGTCAAGGA	840
DB	781	GGAGAACTTGTACATAATCGATTAAGCTTTAAATGATTAAGCTTTGGTCAAGGA	840
QY	841	AGAACAGGTTGGTAGATTTTGGGAAAGTTAAAGGCCCAACATTCGGAATAGCAGGATGCA	900
DB	841	AGAACAGGTTGGTAGATTTTGGGAAAGTTAAAGGCCCAACATTCGGAATAGCAGGATGCA	900
QY	901	GTGAATGCAAGCTAGTCTATTTCTCAGGAAGAGGTTTGGCCACCAATTCGATATAATCGATT	960
DB	901	GTGAATGCAAGCTAGTCTATTTCTCAGGAAGAGGTTTGGCCACCAATTCGATATAATCGATT	960
QY	961	CCTTCAGATCCAGAAAGGAAAGATATATGTCATCCCATACACACCAAGCAGCTCAT	1020
DB	961	CCTTCAGATCCAGAAAGGAAAGATATATGTCATCCCATACACACCAAGCAGCTCAT	1020

QY	1021	TTTGATATAGATTGGGTAAGAAGATGATTCCAATCTCTTAATAGGCATCTATGAATAT	1080
DB	1021	TTTGATATAGATTGGGTAAGAAGATGATTCCAATCTCTTAATAGGCATCTATGAATAT	1080
QY	1081	GGTTATGGCAGTTGGGAAATGATAAAATGGATCTCTCAGTTTGACACAGAAGATT	1140
DB	1081	GGTTATGGCAGTTGGGAAATGATAAAATGGATCTCTCAGTTTGACACAGAAGATT	1140
QY	1141	TTACCTGTAGTCCAGATAAGAAACCCAGGCTTAAGCACTTACAGACTCGTCAGATTAC	1200
DB	1141	TTACCTGTAGTCCAGATAAGAAACCCAGGCTTAAGCACTTACAGACTCGTCAGATTAC	1200
QY	1201	CTCATTAAATTAATCTGAATAAGACCTTGAAGAAGAAAGACACACAGAGCTTGTCTGTGCA	1260
DB	1201	CTCATTAAATTAATCTGAATAAGACCTTGAAGAAGAAAGACACACAGAGCTTGTCTGTGCA	1260
QY	1261	GGCAATTCAGAGGAGAGAAAAACAAGAGTAAGAGATAAAGCAACAAAGGCTGC	1316
DB	1261	GGCAATTCAGAGGAGAGAAAAACAAGAGTAAGAGATAAAGCAACAAAGGCTGC	1316
RESULT 2			
AF181826			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
AUTHORS			
TITLE			
JOURNAL			
FEATURES			
source			
gene			
CDS			



EEERKELEIYMLPRMNCARQISFNENGRCSRSRRYSQSDSISERKPKRRG  
RPTIPRENKGFSDAIEIRRFKSYKKEGGPLERLDAIARDAELVSHKEDLRGEL  
VHNGCTKALNDNDEFGQRTGGRRGVKPTFRFAGVQVNAKLVISHKEELAPLHKSI  
SDPERKRVIPYIYTHKAHFIDWKGEDSNLLIGIYEGYKSWEMIKMDPLSLTK  
ILPDDPKK"

BASE COUNT 1023 a 478 c 663 g 779 t  
ORIGIN

Query Match 87.1%; Score 1146.8; DB 5; Length 2943;  
Best Local Similarity 98.9%; Pred. No. 1.1e-247;  
Matches 1151; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 1 ATTTATCGCTAGTCACAAAGGATCAGTAGAAGAGATATCTTGAAGAGCCAGAAA 60  
DB 1780 ATTTATCGCTAGTCACAAAGGATCAGTAGAAGAGATATCTTGAAGAGCCAGAAA 1839  
QY 61 AAGATGCTGTAGATCAATTTAGTATTCAGAGATCGACACACAGGAAACTGTACTA 120  
DB 1840 AAGATGCTGTAGATCAATTTAGTATTCAGAGATCGACACACAGGAAACTGTACTA 1899  
QY 121 CATACAGGCTTACTCTCAAGCTCAACACCTTTTAATAAGAGAGATTTATCAGCAAT 180  
DB 1900 CATACAGGCTTACTCTCAAGCTCAACACCTTTTAATAAGAGAGATTTATCAGCAAT 1959  
QY 181 TTCAAGTTTGGTGTGAGGAACCTTTTAAAGAACCTGAANNNGAAGAGAGGAGCTCAG 240  
DB 1960 TTCAAGTTTGGTGTGAGGAACCTTTTAAAGAACCTGAANNNGAAGAGAGGAGCTCAG 2019  
QY 241 GAGATGGATATAGATGAATCTCTGAAGAGGCTGAAACTCGAGAAATAGTCCAGGCCCA 300  
DB 2020 GAGATGGATATAGATGAATCTCTGAAGAGGCTGAAACTCGAGAAATAGTCCAGGCCCA 2079  
QY 301 TTAACCTGTAGAGATGAGTTACTTTTACAGTCAAGGTAGCTTAACCTTTTCCAATATGAT 360  
DB 2080 TTAACCTGTAGAGATGAGTTACTTTTACAGTCAAGGTAGCTTAACCTTTTCCAATATGAT 2139  
QY 361 GAAGATGACATTGAATTTGGAACTCAGAACCAAACTTAAGAACTGGGAAGATCAATCCCA 420  
DB 2140 GAAGATGACATTGAATTTGGAACTCAGAACCAAACTTAAGAACTGGGAAGATCAATCCCA 2199  
QY 421 GAAGTTCAGTCCGACAGATAGAGGNGNNGAAGACAAAAGAACTTGAAGAAATATAT 480  
DB 2200 GAAGTTCAGTCCGACAGATAGAGGNGNNGAAGACAAAAGAACTTGAAGAAATATAT 2259  
QY 481 ATGCTTCCAGATGAGAACTGTGCAAAACAGATCAGCTTTTAAATGGAATGAAGGAGA 540  
DB 2260 ATGCTTCCAGATGAGAACTGTGCAAAACAGATCAGCTTTTAAATGGAATGAAGGAGA 2319  
QY 541 TGCAGTACGAGCAGAGATATTTCTGGATCTGATAGTATTCATCTCAGAAAGAAACGA 600  
DB 2320 TGCAGTACGAGCAGAGATATTTCTGGATCTGATAGTATTCATCTCAGAAAGAAACGA 2379  
QY 601 CCAAAAAACGTGGACGACCACTATTCCTCCCGTGAAGAAATTTGGTGGCCCACTTGAAGGTTA 660  
DB 2380 CCAAAAAACGTGGACGACCACTATTCCTCCCGTGAAGAAATTTGGTGGCCCACTTGAAGGTTA 2439  
QY 661 GCAGAGATTAGACGATTTATCAAGAGTTTACAGAAATTTGGTGGCCCACTTGAAGGTTA 720  
DB 2440 GCAGAGATTAGACGATTTATCAAGAGTTTACAGAAATTTGGTGGCCCACTTGAAGGTTA 2499  
QY 721 GATGCTATAGCTAGAGATGCTGAGCTAGTGTATTAATCTGAAGACAGACCTTAGACCTCTG 780  
DB 2500 GATGCTATAGCTAGAGATGCTGAGCTAGTGTATTAATCTGAAGACAGACCTTAGACCTCTG 2559  
QY 781 GGAGAACTGTACATATGATGATCAATTAAGCTTTTAAATGATATGATCTTGGTCAAGA 840  
DB 2560 GGAGAACTGTACATATGATGATCAATTAAGCTTTTAAATGATATGATCTTGGTCAAGA 2619  
QY 841 AGAACAGTGTAGATTTGGGAAGTTAAGGCCCAACATTCAGGAATAGCAGAGTGCAG 900  
DB 2620 AGAACAGTGTAGATTTGGGAAGTTAAGGCCCAACATTCAGGAATAGCAGAGTGCAG 2679

## RESULT 3

A58691  
LOCUS A58691 Sequence 10 from Patent W09639505. DNA linear PAT 06-MAR-1998  
DEFINITION  
ACCESSION A58691  
VERSION A58691.1 GI:3714250  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unclassified.  
REFERENCE 1 (bases 1 to 6608)  
AUTHORS Griffiths R. and Tiwari B.  
TITLE AVIAN GHD GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN BIRDS  
JOURNAL Patent: WO 9639505-A 10 12-DEC-1996;  
COMMENT TSIS INNOVATION (GB)  
FEATURES Other publication AU 5906996 961224.  
location/Qualifiers  
source 1..6608  
/organism="unidentified"  
/db\_xref="taxon:32644"  
BASE COUNT 2289 a 1207 c 1459 g 1653 t  
ORIGIN

Query Match 83.9%; Score 1104.6; DB 6; Length 6608;  
Best Local Similarity 90.4%; Pred. No. 3.2e-238;  
Matches 1176; Conservative 0; Mismatches 125; Indels 0; Gaps 0;

QY 1 ATTTATCGCTAGTCACAAAGGATCAGTAGAAGAGATATCTTGAAGAGCCAGAAA 60  
DB 2943 ATTTATCGCTAGTCACAAAGGATCAGTAGAAGAGATATCTTGAAGAGCCAGAAA 3002  
QY 61 AAGATGCTGTAGATCAATTTAGTATTCAGAGATGGACACACAGGAAACTGTACTA 120  
DB 3003 AAGATGCTGTAGATCAATTTAGTATTCAGAGATGGACACACAGGAAACTGTACTA 3062  
QY 121 CATACAGGCTTACTCTTCAAGCTCAACACCTTTTAAAGAGAGAGTATCAGCAAT 180  
DB 3063 CATACAGGCTTACTCTTCAAGCTCAACACCTTTTAAAGAGAGAGTATCAGCAAT 3122  
QY 181 TTGAAGTTTGGTGTGAGGAACTTTTAAAGAGAGTCTGAANNNGAAGAGAGGAGCCCTAG 240  
DB 3123 TTGAAGTTTGGTGTGAGGAACTTTTAAAGAGAGTCTGAANNNGAAGAGAGGAGCCCTAG 3182  
QY 241 GAGATGGATATAGATGAATCTTGAAGAGGCTGAAGAGGCTGAAGAGGAGGAGCCCA 300  
DB 3183 GAGATGGATATAGATGAATCTTGAAGAGGCTGAAGAGGCTGAAGAGGAGGAGCCCA 3242  
QY 301 TTAAGTGTAGGAGATGAGTTTACTTTTCAAGTTTCAAGTTTCAAGTTTCAAGTTTCAAGTT 360  
DB 3243 TTAAGTGTAGGAGATGAGTTTACTTTTCAAGTTTCAAGTTTCAAGTTTCAAGTTTCAAGTT 3302

[illegible]

Matches 1176; Conservative 0; Mismatches 125; Indels 0; Gaps 0;									
QY	1	ATTATFCGGCTAGTCACAAAGGATCAGTAGAAGAAGATATTCCTTGAAGAGCGCAAGAA	60						
DB	2943	ATTATFCGGCTAGTCACAAAGGATCAGTAGAAGAAGATATTCCTTGAAGAGCGCAAGAA	3002						
QY	61	AAGATGGTGTAGATCATTTAGTGTATTCAGAGATGCACACACAGGAGAAACCTGTACTA	120						
DB	3003	AAGATGGTGTAGATCATTTAGTGTATTCAGAGATGCACACACAGGAGAAACCTGTACTG	3062						
QY	121	CATACAGGCTTACTCCTTCAAGCTCAACACCTTTTAAAGAGAGAGATTCATCAGCAATT	180						
DB	3063	CATACAGGCTTCAACTCCATCAAGCTTACACCTTTTAAAGAGAGAGATTCATCAGCTATT	3122						
QY	181	TTCAAGTTTGGTGTGAGGAACCTTTTAAAGAGAGAGATTCATCAGCAATT	240						
DB	3123	TTCAAGTTTGGTGTGAGGAACCTTTTAAAGAGAGAGATTCATCAGCAATT	3182						
QY	241	GAGATGATATAGATGAATCTCTGAAGAGAGAGATTCATCAGCAATT	300						
DB	3183	GAGATGATATAGATGAATCTCTGAAGAGAGAGATTCATCAGCAATT	3242						
QY	301	TTAACTGTAGAGATGAGTACTTTTCAAGTTCAGGTAGTACCTTTTCCATATATGAT	360						
DB	3243	TTAACTGTAGAGATGAGTACTTTTCAAGTTCAGGTAGTACCTTTTCCATATATGAT	3302						
QY	361	GAGATGATATGAGTACTTTTCAAGTTCAGGTAGTACCTTTTCCATATATGAT	420						
DB	3303	GAGATGATATGAGTACTTTTCAAGTTCAGGTAGTACCTTTTCCATATATGAT	3362						
QY	421	GAGATGATGAGTACTTTTCAAGTTCAGGTAGTACCTTTTCCATATATGAT	480						
DB	3363	GAGATGATGAGTACTTTTCAAGTTCAGGTAGTACCTTTTCCATATATGAT	3422						
QY	481	ATGCTTCCAAAGATGAGAACTGTCAAAACAGATCAGCTTTTAAAGAGAGAGATTCAT	540						
DB	3423	ATGCTTCCAAAGATGAGAACTGTCAAAACAGATCAGCTTTTAAAGAGAGAGATTCAT	3482						
QY	541	TGAGTATGAGAGACAGATATTTCTGATCTGATGATTTTCAAGAGATTTTCAAGAGAT	600						
DB	3483	TGAGTATGAGAGACAGATATTTCTGATCTGATGATTTTCAAGAGATTTTCAAGAGAT	3542						
QY	601	CCAAAGAGTGTGAGAGACAGATATTTCTGATCTGATGATTTTCAAGAGATTTTCAAG	660						
DB	3543	CCAAAGAGTGTGAGAGACAGATATTTCTGATCTGATGATTTTCAAGAGATTTTCAAG	3602						
QY	661	GCAGATATGAGAGACAGATATTTCTGATCTGATGATTTTCAAGAGATTTTCAAGAGAT	720						
DB	3603	GCAGATATGAGAGACAGATATTTCTGATCTGATGATTTTCAAGAGATTTTCAAGAGAT	3662						
QY	721	GATGCTATAGCTAGAGATGCTGAGTATGATGATGATGATGATGATGATGATGATGAT	780						
DB	3663	GATGCTATAGCTAGAGATGCTGAGTATGATGATGATGATGATGATGATGATGATGAT	3722						
QY	781	GGAGAACTTGTACATATGATGATGATGATGATGATGATGATGATGATGATGATGAT	840						
DB	3723	GGAGAACTTGTACATATGATGATGATGATGATGATGATGATGATGATGATGATGAT	3782						
QY	841	AGAGAGGTTGTAGATTTGGGAAAGTTAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAG	900						
DB	3783	AGAGAGGTTGTAGATTTGGGAAAGTTAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAG	3842						
QY	901	GTAAGTGAAGAGTGTATTTCTCAGAGAGAGATGATGATGATGATGATGATGATGAT	960						
DB	3843	GTAAGTGAAGAGTGTATTTCTCAGAGAGAGATGATGATGATGATGATGATGATGAT	3902						
QY	961	CCTTCAGATCAGAGAGAGAGAGAGATGATGATGATGATGATGATGATGATGATGAT	1020						
DB	3903	CCTTCAGATCAGAGAGAGAGAGAGATGATGATGATGATGATGATGATGATGATGAT	3962						
QY	1021	TTTCATATAGATGGGGTAAAGAGATGATGATGATGATGATGATGATGATGATGAT	1080						
DB	3963	TTTCATATAGATGGGGTAAAGAGATGATGATGATGATGATGATGATGATGATGAT	4022						

QY 1081 GGTATGCGAGTGTGGAAATGATAAAATGGATCCTGATCTAGTTTGACACAGAAATTT 1140

DB 4023 GGTATGCGAGTGTGGAAATGATAAAATGGATCCTGATCTAGTTTGACACAGAAATTT 4082

QY 1141 TTACTGATGATCCAGATAAAGAAACCCAGGCTAAGCAGTTACAGACTCGTGCAGATTTAC 1200

DB 4083 TTACTGATGATCCAGATAAAGAAACCCAGGCTAAGCAGTTACAGACTCGTGCAGATTTAC 4142

QY 1201 CTCATTAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1260

DB 4143 CTCATTAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4202

QY 1261 GGCATTTCAAG 1301

DB 4203 GGCATTTCAAG 4243

RESULT 5

AF181824

LOCUS

DEFINITION

ACCESSION

AF181824

VERSION

AF181824.1

GI:5917748

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

20384598

REFERENCE

2

(bases 1 to 2754)

Fridolfsson, A.-K. and Ellegren, H.

AUTHORS

TITLE

JOURNAL

FEATURES

Source

1. 2754

/organism="Aegolius funereus"

/db\_xref="taxon:103375"

/chromosome="W"

<1..>2754

/gene="CHD1W"

<1..>2754

/gene="CHD1W"

/codon\_start=1

/product="chromodomain helicase DNA binding protein 1"

/protein\_id="AAD56021.1"

/db\_xref="GI:5917753"

/translation="ELTDLLKQVQIVRIIAHSNOKSAAAGYDDYCKWGLPYSECS

WEDGALLAKFKQACIDEVFSNOSKTTPEKCKILKORPREVALKKQPSYIGGHESLE

LRDYLQNLNLHSAWCKNSCILLADMGILKGTIOTISFLNLFHEHOLYGPFLVLP

LSTTTSWRIOTIWAPOMNSVYLGDIISRMIRTHEMHPOTKRLKNILLTYEIL

LKDSFGLGNVFIQVDEAHLKNDLSLYKTLIDFKSNHRLITGTPLQNSLKELW

SLLHFMPEKFSWEDFEHKGREFGASLHKEPFLLRVKKVDEKSLPAKVEQ

ILRMESALOKOYKWLIRNYKLSKSGSTSGFLINMELKCCNHCYLKPKPD

NEFYKQEAOLHILRSSGLLILDLRLRGRNVLIFSOHWRMLDILAEYLKRYQ

PFPELDSFKGELKQALDHFNESGDFCLSTRAGGLINLASADTVVIFDSW

PFQNDLQONARRHRCQKKQVNTVRLVTKGSVEEDILERAKKWLDHLVLRMDTG

KIVHTGSTFSSSTPFNKEELSAILFAGAEFLKEPGESEFQEMDILKRAETR

ENEPGLTVGDELLSQIKVNFNSMOEDDIELEPQNSKTSISERKPKRRKRPRTI

QRELEIYMLPMMRCAKQISFNGSGSKRSRKTSGSDSDSISERKPKRRKRPRTI

PREMIKGFDAEIRPKFISYKFGGPEPLERDAIAHDAELVDKSEELAPLHKSIPDNGC

IKALKDNSFQERAGRLGKYPGTFRISGVQVNAKLVISHEELAPLHKSIPDNGC

PKRYVFCHTKAHFDIDWGEDDNLIIYIYEGYGVGEMIKMDPDLSTQKLLPDD

PDKK"

943 a

461 c

623 g

727 t

BASE COUNT

ORIGIN

Query Match 80.1%; Score 1054; DB 5; Length 2754;  
Best Local Similarity 93.9%; Pred. No. 8.2e-227;  
Matches 1093; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

QY 1 ATTATCGGCTAGTACAAAAGGATCAGTAGAAGAGATATCTTCAAAGAGCCAAAGAA 60  
- Db 1591 ATTATCGGCTAGTACAAAAGGATCAGTAGAAGAGATATCTTCAAAGAGCCAAAGAA 1650  
QY 61 AGATGGTGTAGATCATTTAGTAGTACAGAGAAATCGACACACAGGAGAAATCTGTACTA 120  
Db 1651 AGATGGTGTAGATCATTTAGTAGTACAGAGAAATCGACACACAGGAGAAATCTGTACTA 1710  
QY 121 CATACAGGCTCTACTCTTCAAGCTCAACACCTTTTAAAGGAAGAGTATCAGCAAT 180  
Db 1711 CATACAGGCTCTACTCTTCAAGCTCAACACCTTTTAAAGGAAGAGTATCAGCAAT 1770  
QY 181 TTGAAGTTTGGTGCTGAGGAACCTTTTAAAGAACCTTGAANNNGAAGAGGAGCCTCAG 240  
Db 1771 TTGAAGTTTGGTGCTGAGGAACCTTTTAAAGAACCTTGAAGGAGAGAGGAGCCTCAG 1830  
QY 241 GAGATGGATATAGATGAATCTGAAGAGGAGGCTGAACTCGAGAAATGAGTCAGGCCCA 300  
Db 1831 GAAATGGATATAGATGAATCTGAAGAGGAGGCTGAACTCGAGAAATGAGTCAGGCCCG 1890  
QY 301 TTAACHTGTAGGAGATCAGTTACTTTTCCACAGTTCAAGTAGCTAACTTTTCCAATATGAT 360  
Db 1891 CTAACTGTAGGAGATCAGTTACTTTTCCACAGTTCAAGTAGCTAACTTTTCCAATATGAT 1950  
QY 361 GAAGATGACATTTGAATTTGAACCCAGAACAAATCTAAGAACTGGGAGAAATCATTTCCA 420  
Db 1951 GAAGATGACATTTGAATTTGAACCCAGAACAAATTTTAAAGAAATGGGAGAAATCATTTCCA 2010  
QY 421 GAAGTTCACTGGCGAGCATAGAGGGGNGGAAAGACAAAGAACTTGAAGAAATATATAT 480  
Db 2011 GAAGTTCACTGGCGAGCATAGAGGGGNGGAAAGACAAAGAACTTGAAGAAATATATAT 2070  
QY 481 ATGCTTCCAAAGATGAGAACTGTGCAAAACAGATCAGCTTTAATGGAATGAAGGAGA 540  
Db 2071 ATGCTCCCAAGATGAGAACTGTGCAAAACAGATCAGCTTTAATGGAATGAAGGAGA 2130  
QY 541 TGCACTAGGAGCAGAGATATATCTTGATCTGATGTAGTATTCATCTCAGAAAGAAACGA 600  
Db 2131 CGCAGTAGGAGCAGAGATATATCTTGATCTGATGTAGTATTCATCTCAGAAAGAAACGA 2190  
QY 601 CCAAAAAACGTTGGACGACACCAAGTATTTCCCGTGAAACATTAAGAGATTTAGTGAT 560  
Db 2191 CCAAAAAACGTTGGACGACACCAAGTATTTCCCGTGAAACATTAAGAGATTTAGCGAT 2250  
QY 661 GCAGAGATTAGACGATTTATCAAGAGTTTAAAGAAATTTGGTGGCCACGTTGAAAGGTTA 720  
Db 2251 GCAGAGATTAGCGGTTTATCAAGAGTTTAAAGAAATTTGGTGGCCACGTTGAAAGGTTA 2310  
QY 721 GATGCTATAGCTAGAGATGCTGAGCTAGTGTATTAATCTGAACACAGACCTTAGACGCTG 780  
Db 2311 GATGCTATAGCTAGAGATGCTGAGCTAGTGTATTAATCTGAACACAGACCTTAGACGCTG 2370  
QY 781 GGAACTTGTACATATGATGATTAAGGCTTTTAAATGATTAAGCACTTTGGTCAAGGA 840  
Db 2371 GGAACTTGTACATATGATGATTAAGGCTTTTAAAGCACTTTGGTCAAGGA 2430  
QY 841 AGAACACTTGGTAGATTTTGGGAAAGTTTAAAGGCCCAACATTCGGAATAGCAGAGTGCAG 900  
Db 2431 AGNACAGAGGATGAGCTTGGGAAAGTTTAAAGGCCCAACATTCGGAATAGCAGAGTGCAG 2490  
QY 901 GTGAATGCAAGCTAGTCTATTTCTCAGCAAGAGAGTGTGGCACCATTGTGATAAATCGATT 960  
Db 2491 GTGAATGCAAGCTAGTCTATTTCTCAGCAAGAGAGTGTGGCACCATTGTGATAAATCGATT 2550  
QY 961 CCTTCAGATCCAGAAAGAAAGATATGTCATCCCATACACACCAAGAGCTCAT 1020  
|||||

Db 2551 CCTTCAGATCCAGAACAGACAGAAAAGATATGTCATCCCATGCCACACCAAGCAGCTCAT 2610  
QY 1021 TTTGATATAGATTGGGTAAGAAGATGATTCCTCAATCTCTTAATAGGCATCTATGAATAT 1080  
Db 2611 TTTGATATAGATTGGGTAAGAAGATGATTCCTCAATCTCTTAATAGGTATTTATGAATAT 2670  
QY 1081 GGTTATGGCAGTTGGGAAATGATAAAATGGATCCCTGATCTCTCAGTTTGACACAGAAGATT 1140  
Db 2671 GGATATGGCAGCTGGGAAATGATAAAATGGATCCCTGATCTCTCAGTTTGACACAGAAGATT 2730  
QY 1141 TTACTGTGATCCAGATAAGAAA 1164  
Db 2731 TTACTGTGATCCAGATAAGAAA 2754

RESULT 6  
AF181827 2754 bp mRNA linear VRT 07-AUG-2001  
LOCUS Nymphicus hollandicus chromosome W chromodomain helicase DNA  
DEFINITION binding protein 1 (CHDIW) mRNA, partial cds.  
ACCESSION AF181827  
VERSION AF181827.1 GI:5917751  
KEYWORDS  
SOURCE cockatiel.  
ORGANISM Nymphicus hollandicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Archosauria; Aves; Neognathae; Psittaciformes; Cacatuidae;  
Nymphicus.  
REFERENCE 1 (bases 1 to 2754)  
AUTHORS Fridolfsson, A.-K. and Ellegren, H.  
TITLE Molecular evolution of the avian CHD1 genes on the Z and W sex  
chromosomes  
JOURNAL Genetics 155 (4), 1903-1912 (2000)  
MEDLINE 20384598  
REFERENCE 2 (bases 1 to 2754)  
AUTHORS Fridolfsson, A.-K. and Ellegren, H.  
TITLE Direct Submission  
JOURNAL Submitted (30-AUG-1999) Department of Evolutionary Biology,  
Evolutionary Biology Centre, Uppsala University, Norbyvagen 18D,  
Uppsala S-75236, Sweden  
FEATURES  
source  
1..2754  
/organism="Nymphicus hollandicus"  
/strain="Lutino cockatiel"  
/db\_xref="taxon:13180"  
/chromosome="W"  
c1..>2754  
/gene="CHDIW"  
c1..>2754  
/gene="CHDIW"  
/codon\_start=1  
/product="chromodomain helicase DNA binding protein 1"  
/protein\_id="RAD56024.1"  
/db\_xref="GI:5917756"  
/translation="ELTDLHKQVOIVERIIAHSNQKSEAGYPDYCKWQGLPYSECS  
WEDGALIAKFKQACIDIEYFNRSLKSTPDKCKILRQRPVALKQPSYIGGHESLE  
LRDQLGNLNLWLAHSCWNCILADEMLGKTQITISFMNYLFPHHQYGFLLVVP  
LSTLTQWORETQWAPOMNAVYLGDTSRNMRTHEWHMPTKRLKNILLTVEIL  
LKDKSFLGLANVFVIGVDEAHLKNDLSLYKTLIDFKSNHRLITGTPLONSRLWL  
SLLHFMPEKFSWEDFEHKGREFYASLHKEPFLRLRVKDKVEKSLPAKVEQ  
ILRMESALQOYKWIILRYKALSKSGSTSGFLNMMLKCCNHCYLIKPDD  
NEFYKQALQHLIRSSGLILDKLLIRLRGRNVLIFSQVMRLDILAEYLKVRQ  
PFPRDKSGIKGELRKAQDHFNAEGSDFCFLISTRAGGLGINLASADTVVIFOSDW  
NPQDLQAAHARIGOKOVNIVRLVTKGSVEEDILERAKKKMYLDHLVLTQMDTGT  
KTVLHGTSPSSSTPFNKEELSAILKFGAELEKEPEGEPEPEPEPEPEPEPEPEPE  
ENEPPTIVGELLISQFVAFNSNMDEDDIELEPQNLRNNEEIIPEVORRIEER  
QKELEKIYMLPFRNCAKQISFNGSERHRSRRYSGSDSISERRKRRGRRTI  
PRENTKFSDAIRIRFKSYKFGGLERDAIARDAELVDKSETDLRLGELVNGC  
IKALNDSFVQERAAAGRLKVGPTFRISGVQVNAKLVISHEEELAPLHKSPSPPEE  
RKRYVPCHTKAAHFDIDCKEDDSNLLIGIYEYGVGSGWEMIKMDPDLSLQKLPDD  
PDKK"

BASE COUNT 947 a 464 c 613 g 730 t  
ORIGIN

Query Match		79.7%; Score 1049.2; DB 5; Length 2754;
Best Local Similarity		93.6%; Pred. No. 9.8e-226;
Matches 1090; Conservative 0; Mismatches 74; Indels 0; Gaps 0;		
QY	1	ATTATTCGGCTAGTCACAAAAGGATCAGTAGAAGAAGATATCTTTGAAAGAGCGCAAGAAA 60
DB	1591	ATTATTCGGCTAGTCACAAAAGGATCAGTAGAAGAAGATATCTTTGAAAGAGCGCAAGAAA 1650
QY	61	AAGATGGTGTAGATCATTTAGTGAATTCAGAGATGCACACAGGAGAAACTGTACTA 120
DB	1651	AAGATGGTGTAGATCATTTAGTGAATTCAGAGATGCACACAGGAGAAACTGTACTA 1710
QY	121	CATACAGGCTCTACTCTCAAGCTCACACCTTTTAATAAGGAAGATTTATCAGCAATT 180
DB	1711	CATACAGGCTCTACTCTCAAGCTCACACCTTTTAATAAGGAAGATTTATCAGCAATT 1770
QY	181	TTCAAGTTTGGTGTGAGGAGCTTTTAAAGAACCTTGAANNNGAAGAGAGGAGCTCAG 240
DB	1771	TTCAAGTTTGGTGTGAGGAGCTTTTAAAGAACCTTGAANNNGAAGAGAGGAGCTCAG 1830
QY	241	GAGATGGATATAGATGAATCTCTGAAGAGNCTGAAACTCGAGAAATAGTCAAGCCCA 300
DB	1831	GAAATGGATATAGATGAATCTCTGAAGAGGCTGAAACTCGAGAAATAGTCAAGCCCA 1890
QY	301	TTAACTGTAGGAGATGATTTTACAGAGTTCAAGGTAGCTAACTTTTCCATATGGAT 360
DB	1891	ATACTGTAGGAGGAGGAATGCTTTACAGAGTTCAAGGTGCTAACTTTTCCATATGGAT 1950
QY	361	GAAGATGACATTGAATTTGAACACAGAACAAATCTTAAGAACTGGGAAGAAATCATCCA 420
DB	1951	GAAGATGACATTGAATTTGAACACAGAACAAATTTAAGAACTGGGAAGAAATCATCCA 2010
QY	421	GAAGTTTCAGTGGCAGCAATAGAGGNGGAAAGACAAAGAACTTGAAGAAATCATCCA 480
DB	2011	GAAGTTTCAGTGGCAGCAATAGAGGNGGAAAGACAAAGAACTTGAAGAAATCATCCA 2070
QY	481	ATGCTTCAAGATGAGAACTGTCAAAACAGATCAGCTTTATGGAATGAGGGAGA 540
DB	2071	ATGCTTCAAGATGAGAACTGTCAAAACAGATCAGCTTTATGGAATGAGGGAGA 2130
QY	541	TGCAATGAGCAGAGATATTCGGATCTGATGATGATTCATCTCAGAAAGAAACGA 600
DB	2131	CACATGAGGAGCAGAGATATTCGGATCTGATGATGATTCATCTCAGAAAGAAACGA 2190
QY	601	CAAAAAACGTGGACGACCACTATTCCTCCGTGAAACATTAAGAGATTAGTGTAT 660
DB	2191	CAAAAAACGTGGACGACCACTATTCCTCCGTGAAACATTAAGAGATTAGTGTAT 2250
QY	661	GCAGAGATTAGCAGATTATCAAGAGTTTCAAGAAATTTGGTGGCCCACTTGAAGGTTA 720
DB	2251	GCAGAGATTAGCAGATTATCAAGAGTTTCAAGAAATTTGGTGGCCCACTTGAAGGTTA 2310
QY	721	GATGCTATAGCTAGAGATCTGAGTGTAGTGAATCTGAAACAGACCTTAGAGCTGT 780
DB	2311	GATGCTATAGCTAGAGATCTGAGTGTAGTGAATCTGAAACAGACCTTAGAGCTGT 2370
QY	781	GGAGACTTGTACATATGATGATGATTAAGGCTTTAAATGATATGATTTGTCACAGA 840
DB	2371	GGAGACTTGTACATATGATGATGATTAAGGCTTTAAATGATATGATTTGTCACAGA 2430
QY	841	AGACAGTGTAGATTGGGAAGTTTAAAGGCCCAACATTCGAATAGCAGAGTGCAG 900
DB	2431	AGACAGTGTAGATTGGGAAGTTTAAAGGCCCAACATTCGAATAGCAGAGTGCAG 2490
QY	901	GTGAATGCAAGCTAGTCAATTTCTCAGAGAAAGAGTGGCCACCATTCGATTAATCGATT 960
DB	2491	GTGAATGCAAGCTAGTCAATTTCTCAGAGAAAGAGTGGCCACCATTCGATTAATCGATT 2550
QY	961	CCTTCAGATCCAGAGAAAGAGATATGTCTATCCCATACACACCAAGAGCAGCTAT 1020
DB	2551	CCTTCAGATCCAGAGAAAGAGATATGTCTATCCCATACACACCAAGAGCAGCTAT 2610
QY	1021	TTTGATATAGATTGGGTAAAGAGATGATTCATCTGTTAATAGGATCTATGATAT 1080
DB	2611	TTTGATATAGATTGGGTAAAGAGATGATTCATCTGTTAATAGGATCTATGATAT 2670
QY	1081	GGTTATGGCAGTTGGGAAATGATAAAATGGATCTGATCTCATGTTGACACAGAAATT 1140
DB	2671	GGCTATGGCAGCTGGGAAATGATAAAATGGACCTGATCTCATGTTGACACAGAAATT 2730
QY	1141	TTACTGTGATCCAGATAAGAAA 1164
DB	2731	TTACTGTGATCCAGATAAGAAA 2754
RESULT 7		
LOCUS	DI4316	2292 bp mRNA linear VRT 03-FEB-1999
DEFINITION	delta-crystallin enhancer binding protein, complete sequence.	
ACCESSION	DI4316	
VERSION	DI4316.1	GI:391639
KEYWORDS	ORF2.	
SOURCE	Gallus gallus (library: lambda gt11) 13 day embryo lens cDNA to mRNA, clone JF11.	
ORGANISM	Gallus gallus	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae;	
AUTHORS	Funahashi, J.	
TITLE	Direct Submission	
JOURNAL	Submitted (29-JAN-1993) Jun-ichi Funahashi, Inst. Devel. Aging and Cancer, Tohoku Univ.; 4-1 Seiryomachi, Aoba-ku, Sendai 980-77, Japan (Tel:022-272-9499, Fax:022-272-3982)	
REFERENCE	2 (sites)	
AUTHORS	Funahashi, J., Sekido, R., Murai, K., Kamachi, Y. and Kondoh, H.	
TITLE	Delta-crystallin enhancer binding protein delta E1 is a zinc finger-homeodomain protein implicated in postgastrulation embryogenesis	
JOURNAL	Development 119 (2), 433-446 (1993)	
MEDLINE	94116444	
REFERENCE	3 (bases 1 to 2292)	
AUTHORS	Funahashi, J.	
JOURNAL	Unpublished (1994)	
FEATURES	Location/Qualifiers	
source	1..2292	
	/organism="Gallus gallus"	
	/db_xref="taxon:9031"	
	/tissue_type="lens"	
	/clone_lib="lambda gt11"	
	/dev_stage="13 day embryo"	
	257..1939	
CDS	/note="beta-gal fusion protein binds to blocks 10 to 3 of HN fragment of delta-crystallin enhancer."	
	/codon_start=1	
	/product="ORF2"	
	/protein_id="BA00262.1"	
	/db_xref="GI:391640"	
	/translation="MELAKCCNHCYLKPPDDNFEYKNQERLQHLIRSGKLILLDKL LIRERGNRLVFSQWMDLILAEYLKRFQPKLDGSLNGELRQALDHFNAEG SEDFCFLSTRAGGLINLADTVIFSDNPNQDLQAAHRIQKQKQVNIYHL VTGSEVEDILERAKKNVLDHLVTQMTTKTVLHTGTPSSSTPPNKELSAILK FGAEELFPEGEQEPQEMDIDELIKRAETRENPGLTVGDELLSQFKVANSND EDDLELEPNRSNWEELIIPESORREKEELIYMLPRMNCARAKQISFNGE GRRSRRYSGSDSITERRKKRKRPTIPRENKIGFSDAEIRRRIRKSKYKFGP LERLDVADAEVLVDKSELDLRLGLHNGICAKLDSSGOERAGGRKGVKGTFF RISGVQNAKLVSHEEELAPLHKISIPDPERKRYVIPCCKAAHFDIDMGKDDSN LLVGIEYAYGWSQCKVNGSRQNLTEILPDPDPTPRQNSRPVQTTSLN"	
BASE COUNT	827 a	385 c 546 g 534 t
ORIGIN		
Query Match	78.3%; Score 1030.4; DB 5; Length 2292;	
Best Local Similarity	88.7%; Pred. No. 1.7e-221;	

Matches 1157; Conservative 0; Mismatches 142; Indels 5; Gaps 4;				
QY	1	ATTATCGGCTAGTCACAAAAGGATCAGTAGAAGAAGATATTCCTGAAAGAGCCAAAGAAA	60	
Db	725	ATTATCGGCTAGTCACAAAAGGATCAGTAGAAGAAGATATTCCTGAAAGAGCCAAAGAA	784	
QY	61	AGATGGTGTAGATCATTTAGTGTATCAGAGAAATGGACACACAGGGAAGAAATGTACTA	120	
Db	785	AGATGGTGTAGATCATTTAGTGTATCAGAGAAATGGACACACAGGGAAGAAATGTACTG	844	
QY	121	CATACAGGCTCTACTCCTCAAGCTCAACACCTTTTAATAAGAAAGAGTTATCAGCAAT	180	
Db	845	CATACAGGTTCAACTCCATCAAGCTCTACACCTTTTAATAAGAAAGAGTTATCAGCTAT	904	
QY	181	TTGAAGTTTGGTGCTGAGGAACCTTTTAAAGAACCTTGAANNNGAAGAAGAGGAGCCTCAG	240	
Db	905	TTGAAGTTTGGTGCTGAGGAACCTTTTAAAGAACCTTGAAGGAGAAGACAGGAGCCCCAG	964	
QY	241	GAGATGGATATAGATGAATCTGAAGAGGNCGTGAACCTCGAGAGAAATGAGTCAGGCCCA	300	
Db	965	GAAATGGATATAGATGAATCTTGAAGAGAGCTGAAACTCGGAGAAATGAGCCAGGTCCA	1024	
QY	301	TTAACTGTAGAGATCAGTTTACTTTCCACAGTTCAGAGTAGCTAACTTTTCCAAATATGGAT	360	
Db	1025	TTCACTGTAGGGATGAGTTGCTTTCCACAGTTCAGAGTGCAGAACTTTTCCAAATATGGAT	1084	
QY	361	GAAGATGACATTTGAATGGAACACAGAACAAATCTAAGAAACTGGGAAGAAATCATTTCCA	420	
Db	1085	GAAGATGATATTCAGTTTGAACCCAGAAAGAAATTCAGAAATTTGGGAAGAAATCATCCCA	1144	
QY	421	GAAGTTCACTGGCGAGATAGAGGNGNGGAAGACACAAAAGAACTTGAAGAAATATAT	480	
Db	1145	GAATCCCAACGGAGAGGATAGAGGAGGAGGAAGAACACAAAAGAACTTGAAGAAATATAT	1204	
QY	481	ATGCTTCCAAAGATGAGAACTGTCAAAACAGATCAGCTTTAATGGAATGAAGGAGA	540	
Db	1205	ATGCTCCGAGGATGAGAACTGTCAAAACAGATCAGCTTTAATGGAATGAAGGAGA	1264	
QY	541	TGCAGTAGGACGAGAGATATTCGTGATGTAGTATTCATTCATTCAGAAAGAAACGA	600	
Db	1265	CGCAGTAGGACGAGAGATATTCGTGATGTAGTATTCATTCATTCAGAAAGAAACGG	1324	
QY	601	CCAAAACAGCTGGACGACCCAGCAACTATTCCTCGTGAACAACTTAAAGAGTTAGTAT	660	
Db	1325	CCAAAACAGCTGGAGAGCCTCGAACCTATTCCTCGAGAAATATTAAGAGTTAGTAT	1384	
QY	661	GCAGAGATTAGAGATTTATCAAGAGTTTACAGAAATTTGGTGCCCACTTGAAGAGTTA	720	
Db	1385	GCAGAGATCAGCGGTTTATCAAGAGTTTACAGAAATTTGGTGCCCACTTGAAGAGTTA	1444	
QY	721	GATGCTATAGCTAGAGATGCTGAGCTAGTGTATTAATCTGAAACAGACCTTAGACCTCTG	780	
Db	1445	GATGCTGTAGCTAGAGATGCTGAACTGGTTGATTAATCTGAGACAGACCTTAGACCTTG	1504	
QY	781	GGAGAACTTGTACATAATGGATGCATTAAGGCTTTTAAATGATAATGACTTTGGTCAAGA	840	
Db	1505	GGTGAAGCTTGTACATAATGGATGCATTAAGGCTTTTAAAGGACAAATTCATCTGACAA	1564	
QY	841	AGAACAGGTGGTATTTGGGAAAGTTTAAAGGCCCAACATTCGGAATAGCAGAGTGCAG	900	
Db	1565	AGACGAGGAGGTAGACTTGGGAAGTTTAAAGGCCCAACATTCGGAATAGCAGAGTGCAG	1624	
QY	901	GTGAATGCAAGAGTACTATTTCTACGAGAGAGAGTGTGGACCAATTCGATTAATATCGAT	960	
Db	1625	GTGAATGCAAACTAGTCTATCTCAGGAAGAGAGTGTGGACCACTGACCAACATTCAT	1684	
QY	961	CCTTCAGATCCAGAAAGGAAAGATATGTCATCCCATACACACCAAGACCTCAT	1020	
Db	1685	CCTTCAGATCCAGAAAGGAAAGATATGTCATCCCATCCATCCATCCACACCAAGGCTG	1744	
QY	1021	TTTGATATAGATTGGGGTAAAGAAAGATGATTCCTCAATCTGTTAATAGGCATCTATGAAT	1080	
Db	1745	TTTCGATATAGATTGGGGTAAAGAGATGATTCCTCAATCTGTTAGTAGGCATCTATGAAT	1804	

BASE COUNT

931 a 481 c 636 g 706 t

QY	1081	GTTATGGCAGTTGGAA-ATGATAAAATGGATCCTGATCTCAGTTTGACACAGAAGAT	1139	
Db	1805	GCCTATGGCAGCTGGCAATGAAGTAAATGGATCCAGATCTCAGCTTAACACAG-AGAT	1863	
QY	1140	TTTACCTGATGATCCAGATAAGAAACCCAGGCTAAGCAGTTTACAGACTGTCGACGATTA	1199	
Db	1864	TTTACCTGATGATCCAGACA-GAACTCCAAGGCAAAACAGCTACAGACCGTCGACACTA	1922	
QY	1200	CCTCATTAATTAATCAATAAAGACCTTCAAGAAAGGAG--CACAGAGACTTGCTGCT	1257	
Db	1923	CCTCATTAATTAATCAATAAAGACCTTCAAGAAAGGAGTGCCTAAAGGCTTGCTGCT	1982	
QY	1258	GCAGGCAATTCGAAGAGGAGAAAAACAAGAAGTAAAGAAGATAA	1301	
Db	1983	GCAGGCAATTCGAAGAGGAGGAGACAAGAAGTAAGAAGATAA	2026	
RESULT	8			
LOCUS	AF181825	2754 bp	mRNA	linear
DEFINITION	Aegolius funereus chromosome 2 chromodomain helicase DNA binding protein 1 (CHD1Z) mRNA, partial cds.			
ACCESSION	AF181825			
VERSION	AF181825.1	GI:5917749		
KEYWORDS	Aegolius funereus.			
SOURCE	Aegolius funereus.			
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Strigiformes; Strigidae; Aegolius.			
REFERENCE	Fridolfsson, A.K. and Ellegren, H.			
AUTHORS	Molecular evolution of the avian CHD1 genes on the Z and W sex chromosomes			
TITLE	Genetics 155 (4), 1903-1912 (2000)			
JOURNAL	20384598			
MEDLINE	2 (bases 1 to 2754)			
AUTHORS	Fridolfsson, A.-K. and Ellegren, H.			
TITLE	Direct Submission			
JOURNAL	Submitted (30-AUG-1999) Department of Evolutionary Biology, Evolutionary Biology Centre, Uppsala University, Norbyvagen 18D, Uppsala S-75236, Sweden			
FEATURES	Location/Qualifiers			
source	1..2754			
gene	/organism="Aegolius funereus"			
CDS	/db_xref="taxon:103375"			
	/chromosome="Z"			
	<1..>2754			
	/gene="CHD1Z"			
	<1..>2754			
	/codon_start=1			
	/product="chromodomain helicase DNA binding protein 1"			
	/protein_id="AAD56022.1"			
	/db_xref="GI:5917754"			
	/translation="ELTDDLHKQYIVERIIAHSNKSAAAGYPDYCKWQGLPYSECS			
	WEDGALIAKFTQCIDSEYFSNRKSTTFKDCVKLQKQRFVAKKQPSYIGGHEGLE			
	LDYOLGNLWLAHSCNKGSCILADEMLGKTIOTISFLNLYLPHFHOLYGPFLLVWP			
	LSTLTWSOREIQTWAPOMNAVYLGDTSRNMIRTHEWHLQTKLKNILLTYYEIL			
	LKDSFLGLNWFAGFYDEAHLKNDSDLLYKTLIDFKSNHRLITGTPQLNSKELW			
	SLLHFIPEKFSWEDFEHKGREYGYSLHKEPFLLRVRKDKVEKSLPAKVEQ			
	ILRMMSALQOYKWIILRYKALSKGSGSTGLNIMKELKCCCHCYLIKPPDG			
	PFYFDLQSGIKGELRKQALDHFNAEGSDFCFLSTRAGLGINLASADTVVIFDSW			
	NEFYNKQALQYIKWILRLKLLRERGNRLIFSOMVRLMDILAEYLRQ			
	NPQDLQAAHRTGOKQVNIYLVTKGVEEDILERAQKWLHLVLTQMDTGT			
	KTVLHGTSPSSPFPNKEELSAILKFAABELFKPEGEQEPQMDIDELKRAETR			
	ENEPGLTVGDELLSQFVANSNDEDDIELEPNSRNWDEISIPVQRRRIEER			
	QKELEIYMLPMRNCARQISFNGSESRSSRYSGSDSISIRKPKKRGRTPI			
	PRENKFSDAEIRRFIKSYKFGPLERDAVDALVDKSETDLRLGLGVNGC			
	IKALKDSSGOERAGRLGKVPTRISRGVQVNAKLVIHSEELAPLHKSTPSPDEE			
	RKRVVPCHTKRAHFDIDWGEDSDNLLIGIYEYGVGSGWEMIKMDPDLSLTQKILPDD			
	PKR"			

## ORIGIN

Query Match 77.5%; Score 1020.4; DB 5; Length 2754;  
Best Local Similarity 92.1%; Pred. No. 2.9e-219; Indels 0; Gaps 0;  
Matches 1072; Conservative 0; Mismatches 52;

Qy 1 ATTATCGGCTAGTCAAAAGAGTACAGTGAAGAGATATCTTGAAGACCCAGAAA 60  
Db 1591 ATTATCGGCTAGTCAAAAGAGTACAGTGAAGAGATATCTTGAAGACCCAGAAA 1650

Qy 61 AGATGGTGTAGATCATTAGTATGATTCAGAGATGGACACACAGGGAATACTGTACTA 120  
Db 1651 AGATGGTGTAGATCATTAGTATGATTCAGAGATGGACACACAGGGAATACTGTACTA 1710

Qy 121 CATACAGGCTCTACTCTTCAAGCTCAACACCTTTTAATAAGGAAGAGTTATCAGCAATT 180  
Db 1711 CATACAGGCTCTACTCTTCAAGCTCAACACCTTTTAATAAGGAAGAGTTATCAGCAATT 1770

Qy 181 TTGAAGTTTGTGTGAGGAACTTTTAAAGAACCTTGAANNNGAAGAGAGGCTCAG 240  
Db 1771 CTGAAGTTTGTGTGAGGAACTTTTAAAGAACCTTGAANNNGAAGAGAGGCTCAG 1830

Qy 241 GAGATGGATATAGATGAATCTGAAGGNETGAACCTCGAAGAAATGAGTCAGGCCCA 300  
Db 1831 GAATGGATATAGATGAATCTGAAGGAGCTGAACCTCGAAGAAATGAGTCAGGCCCA 1890

Qy 301 TTAACCTGAGAGATGAGTACTTTCACAGTTCAGGTAGTACTTTCCTCAATATGAT 360  
Db 1891 TTAACCTGAGAGATGAGTACTTTCACAGTTCAGGTAGTACTTTCCTCAATATGAT 1950

Qy 361 GAAGATGACATGAATGGAAACCAAGAAATCTAAGAACTGGGAAGAAATCATTCCTCA 420  
Db 1951 GAAGATGACATGAATGGAAACCAAGAAATCTAAGAACTGGGAAGAAATCATTCCTCA 2010

Qy 421 GAAGTTCAGTGGCGACGATAGAGGNGGGAAGACAAAGAACTTGAAGAAATATAT 480  
Db 2011 GAAGTTCAGTGGCGACGATAGAGGNGGGAAGACAAAGAACTTGAAGAAATATAT 2070

Qy 481 ATGCTTCCAGATGAGAACTGTGCAAAACAGATCAGCTTTAATGAAATGAAGGAGA 540  
Db 2071 ATGCTTCCAGATGAGAACTGTGCAAAACAGATCAGCTTTAATGAAATGAAGGAGA 2130

Qy 541 TGCAGTAGGACGAGATATCTTGGATCTGATGATGATTCATCTCAGAAAGAAACGA 600  
Db 2131 CGCAGTAGGACGAGATATCTTGGATCTGATGATGATTCATCTCAGAAAGAAACGA 2190

Qy 601 CCAAAAACGTGGACGACCGACACTATTCCTCGTGAACATTAAGGATTTAGTAT 660  
Db 2191 CCAAAAACGTGGACGACCGACACTATTCCTCGTGAACATTAAGGATTTAGTAT 2250

Qy 661 GCAGAGATTAGACGATTTATCAAGATTTCAAGAAATTTGTGGCCAGTGAAGGTTA 720  
Db 2251 GCAGAGATTAGACGATTTATCAAGATTTCAAGAAATTTGTGGCCAGTGAAGGTTA 2310

Qy 721 GATGCTATAGTATGATGCTGAGTATGATTAATCTGAAACAGACCTTAGACGCTG 780  
Db 2311 GATGCTATAGTATGATGCTGAGTATGATTAATCTGAAACAGACCTTAGACGCTG 2370

Qy 781 GGAGAACTTGTACATAATGATGATTAAGGCTTTAATATGATTAATGATTTGGTCAAGGA 840  
Db 2371 GGAGAACTTGTACATAATGATGATTAAGGCTTTAATATGATTAATGATTTGGTCAAGGA 2430

Qy 841 AGAAGAGTGTAGATTGGGAAGTTTAAAGGCCCAACATTCGAATATGAGGAGTGCG 900  
Db 2431 AGAAGAGTGTAGATTGGGAAGTTTAAAGGCCCAACATTCGAATATGAGGAGTGCG 2490

Qy 901 GTGAATCAAAAGCTTAGTCAATTTCTCAGAGAGAGTTGGACCATTCGATTAATCGATT 960  
Db 2491 GTGAATCAAAAGCTTAGTCAATTTCTCAGAGAGAGTTGGACCATTCGATTAATCGATT 2550

Qy 961 CCTTCAGATCAGAGAAAGAGATATGTCATCCCATACACACCAAGGAGCTCAT 1020  
Db 961 CCTTCAGATCAGAGAAAGAGATATGTCATCCCATACACACCAAGGAGCTCAT 1020

Db 2551 CCATCAGATCCAGAAAGAAAGATAGTACGTATCCATCCATCCACACCAAGGCTGCTCAT 2610

Qy 1021 TTTGATATAGATTTGGGTAAAGAGATGATTCCTCAATCTGTTAATAGGATCTATGATAT 1080

Db 2611 TTTGATATAGATTTGGGTAAAGAGATGATTCCTCAATCTGTTAATAGGATCTATGATAT 2670

Qy 1081 GGTATGGCAGTGGGAAATGATAAAATGGATCCTGATCTCAGTTTGACACAGAGATT 1140

Db 2671 GGTATGGCAGTGGGAAATGATAAAATGGATCCTGATCTCAGTTTGACACAGAGATC 2730

Qy 1141 TTACTCTGATCCAGATAAGAAA 1164

Db 2731 TTGCTGATCCAGATAAGAAA 2754

RESULT 9

AF181828 2754 bp mRNA linear VRT 07-AUG-2001

LOCUS Nymphicus hollandicus chromosome 2 chromodomain helicase DNA binding protein 1 (CHD12) mRNA, partial cds.

DEFINITION Nymphicus hollandicus chromosome 2 chromodomain helicase DNA binding protein 1 (CHD12) mRNA, partial cds.

ACCESSION AF181828

VERSION AF181828.1 GI:5917752

KEYWORDS

SOURCE cockatiel.

ORGANISM Nymphicus hollandicus

REFERENCE 1 (bases 1 to 2754) Frilofsson, A.K. and Ellegren, H. Molecular evolution of the avian CHD1 genes on the Z and W sex chromosomes. Genetics 155 (4), 1903-1912 (2000)

REFERENCE 2 (bases 1 to 2754) Frilofsson, A.K. and Ellegren, H. Direct Submission

AUTHORS Submitted (30-AUG-1999) Department of Evolutionary Biology, Evolutionary Biology Centre, Uppsala University, Norbyvägen 18B, Uppsala S-75236, Sweden

JOURNAL Uppsala S-75236, Sweden

TITLE Location/Qualifiers

FEATURES

source 1. 2754

organism="Nymphicus hollandicus"

strain="Lutino cockatiel"

db\_xref="taxon:13180"

chromosome="2"

cl. .>2754

gene="CHD12"

gene="CHD12"

codon\_start=1

product="chromodomain helicase DNA binding protein 1"

protein\_id="AAD56025.1"

db\_xref="GI:5917752"

translation="ELTDLLHQYQIVERTIAHSNOKSAAGYPDHYCKWQGLPYSECS WEDGALIAKFOARIDEVFSRNSQSTTPPKCKVKLQKORFVALKKQPLYIGGHESLE LRDYQNLGNLWASHCKNSGCIADENGLGKTIQTISFLNLFHEHOLYGPFLVVP LSTLTSQREIQTWASQMNAYVIGDITSRNMIRTHEMHPQT KLFNILLTYYEIL LKDKSFLGGLNWFATGVDEAHLKNDLSLLYKTLIDFKSNHRLITGTPQNLSKELW SLHFTMPKFSWEGPEEHGKREYGVASLHKELPFLLRVKDKVKSIPAKVEO ILRMESALOKOYKWIITRYNKAISKSGSTSGFLNMELKCCNHCYLIKPPDO NEFYKOEALQHLIRSSKLLIDKLLIRLREGRNRYLIFSOMVRMLDILAEYLKPYRO PPNLDGSGIKGELKQALDHFNAEGSEDFCLPILLSRAGGLINLASADTVVIFDSWM PNPDLOQAARHRTGOKOVNIYRLVTKGVSVEEDILERAKKKMWLDHLVLRMDTTC KTVLHGTSTPSSSTPNKEELSAILKFAEELFKPEEGEPQEMDIDELIKFAETR ENEGPLTVGDILLSQKVAENFSNDEDDILEPERNCRNWEIIPVQORREIPEER OKELBEITYMLPRMNAKQISFNGSGRRSRYSRSGSDSISERKKPKRGKGPRTI PRENKGFDSEAIRFTKSKKFGGFLERLDVARDALVDKSEIDRUGELVHNGC IKAALDSSQDERAGRGKGTFRISGVGVNAKLVISSHEELAPLHKSPIDPDEE RKRIVIPCHTKAAHFDIDWKGEDSNLLIGIYEYIGSWEMIKMDPDLSLTKILPDD PDKK"

BASE COUNT 926 a 642 g 716 t

ORIGIN







3728	CGACAGGTGGTAGACTCGGAAAGTGAAGGGTCCACATTCCTCGAATATCAGGAGTACAG	3787
901	GTGATGCGAAAGCTAGTCTATTTCTCAGGAAGAAGATTTGGCACCATTTCGATAAATCGATT	960
3788	GTGATGCGCAATAGTCAATCTCCCATGAAGAAGAATTAATATCATCTTGCACAAMATCCATT	3847
961	CGTTCAGTCAACAGCAAGAAGCAAGATATGTCTACCTCATCAGACCAACCAAGACGTCATT	1020
3848	CGTTCTGATCTCCAGACAGAAGAAGACATTAATCTATCTCCATCTCCACACAAGGCGAGCTCAT	3907

[illegible]

RESULT 11	
MUSCULIX	
LOCUS	MUSCHDLX 5349 bp mRNA linear ROD 19-SEP-1996
DEFINITION	Mouse DNA-binding protein (CHD-1) mRNA, complete cds.
ACCESSION	L10410.X66028
VERSION	L10410.1 GI:455014
KEYWORDS	DNA binding protein.
SOURCE	house mouse.
ORGANISM	Mus musculus.
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
AUTHORS	1 (bases 1 to 5349) Delmas, V., Stokes, D.G. and Perry, R.P.
TITLE	A mammalian DNA-binding protein that contains a chromodomain and an SNF2/GWR2-like helicase domain
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 90 (6), 2414-2418 (1993)
MEDLINE	93211972
REFERENCE	2 (bases 1 to 5349) Perry, R.P.
AUTHORS	Direct Submission Perry, R.P.
TITLE	Submitted (08-APR-1993) Robert Perry, The Fox Chase Cancer Center, 7701 Burholme Avenue, Philadelphia, PA 19111, USA
JOURNAL	On Feb 16, 1994 this sequence version replaced gi:2933322.
COMMENT	

```

source
1..5349
/organism="Mus musculus"
/db_xref="taxon:10090"
/cell_type="plasmacytoma"
1..5349
/gene="CHD-1"
171..5306
/gene="CHD-1"
/codon_start=1
/product="RNA-binding protein"
/protein_id="A008486.1"

```

```

/unicode="01433013"
/translation="MNGHSDSESVRNGSGSGSDGCGSASGSGSGSGSGSDSGS
SQSGSDSGSDSGSDSGSDSTRENNYQAPKPKVDGAEFWKSPS1LAVQRSML
RQPOQAOQRPASNSGSEEDSDSDSGARKKHNDWQMGSGSPQALG
SUSEEERDKSCDGTESDYEPPKNYRSRKPNRSKSKNGKKL1LQKKRQIDSEDE
DEEDYDNDKRSRQ1TYAVDQSDMLLEVCEDGDPQDEEFT1ERVV
DCVRGKATGAT1TYAVDQSNAGPBRNNKEPDQIYL1KWKGW1HNTWETEE

```

TLKQONVRGMKKLDNKKKQDETNRWLNKASPEDVEYNNCOQLTDDLLHKQYQIVERI  
TAHNSOKSAAGLPDYKQWQGLPYSECWEDGALISKKFFQTCIDIFYRNOSKTPPFK  
DCKVLKQPRFVALKQPSYIGGHEGLERLDYQNLNLWLAHSWCKNSCILAEDMGL  
GKTIOITISFLNLPHEHOLYGFLLVPLSLTWSOREIOIOTWASOMNVAHVYLIDNSR  
NMIRTHMHPOTKRLKNILLTVEILLKDKAFTGLGNWAFIGVDEARHLKNDLSL  
YKTLIDFSNHRLLITGTPLONSLXELSLHLHFMPEKFSWEDFEHGGKRGYGA  
SLHKEFLPLRRVKDEKSLPAKVEQLRMESALQKQYKWLITLRYKALSXGSK  
GSTGFLNIMMELKCCNCHYLKPPDNEFYKQELHLSRSGKLLILDKLLIRL  
REGRNVLIFSQWMLDILAEYKYRFPQRLDGSIKGELRKQALDHFNAEGSEDF  
CFLSTRAGGLGINLASADTVIFDSWPNPONDLOAQAHRIGOKKOVNIVRLVTKG  
SVEEDILERAKKWVLDHLVIOEMDTCKTVLHTGSAOSSPTPENKEELSAILKEGAE  
ELFKPEGEQEPQMDIDEILKRAETHENEPGLSVGDELLSQKYVANFNMDDEDDI  
ELEPERNKWEEIIPERORRKEEEROKLEIYMLPRMNCQAISFNSEGBERR  
DSRTSGSDSISSRKRPKRGPRPTIPRENKIGFSDAETRRFTKSKYKFGGLPRIS  
DAIARDAELVKSFTDLRLGELVHNGCVKALKDSSGTERAGGLVKVGPPTFRISG  
VOVNAIWAHELDILPLHKSIPSPDEKQVTPICHTKAAHFDIDWKEDDLSNLLIG  
IYEXGYGSMWIMKDPDLSLTHKILPDPDKKPOAKOLOTRADYLILKNDLAKREA  
ORLCAGAGSKRRKTRAKSKAMKSIVKEEIKSDSPSPSEKSDDDDKLNDSPESK  
DRSKSVSDAPVHITASGEVPVIAESEEEDQKTFISICKERMVRPVAALKOLDREPK  
GLSREQLHTKIKIGDHTECLYSNPEQIKOWRKNLWIFVSEFTEDARKHL  
KLYKHAIKROESQONSQNSVATYHVRNPDMERLKENYHNDSDRSYSSDRSL  
QYDHHKROHGDYSKSDSRKRPYSFNSCKDHEWDHYRODSRYYSDRKHKRLDD  
HRSRHSPLSGLKDRCHSDHSHDRMSDHSSEHHTHKSRSRYRILSDWQLD  
HRAASGPRSPDLQKSPYSGSRSPFHSAEHRSPTPEHTWSSRKT\*

BASE COUNT 1739 a 1067 c 1319 g 1224 t

Query Match 67.2%; Score 884.6; DB 10; Length 5349;  
Best Local Similarity 79.4%; Pred. No. 9.1e-189;  
Matches 1043; Conservative 0; Mismatches 270; Indels 0; Gaps 0;

QY 1 ATTATCGCTAGTACACAAAGAGCAGTAGAAGAGATATCTTGAAGAGCGCAAGAA 60 10  
DB 2889 ATATATCGCTGGTTACAAGGGATCAGTTGAAGAGATATCTTGAAGGGCCCAAG 2948

QY 61 AGAGTGGTGTATGATCATTTAGTATCAGAGAATGGACACACAGGGAATACTGTACTA 120  
DB 2949 AAAATGGTTTGGATCATCTTGTATCAAGAATGGATACCACTGGGAAGACAGTGC 3008

QY 121 CATACAGCTCTACTCTTCACAGCTCAACACCTTTTAAATGAAGAGAGTTATCAGCAAT 180  
DB 3009 CACACAGGCTCGGCTCCGTCAGTCCACCCCTTCAATAAAGAGAGGATTATCCGCCAT 3068

QY 181 TTGAAGTTTGGTCTGAGGAACCTTTTAAAGAACCTTGAANNNGAAGAGAGGCTCAG 240  
DB 3069 TTAAATTCGGTCTGAGGAGCTCTTAAGAGGCTTGAAGAGAGGAGGAGGAGCCACAG 3128

QY 241 GAGATGGATATAGTAAATCTGAAGAGGNGCTGAAACTCGAGAAATGAGTCAGGCCCA 300  
DB 3129 GAAATGGATATCGATGAAATCTGAGAGAGCGGAGACTTCAGGAGAACGAGCGGCCCA 3188

QY 301 TTAAGTGTAGGAGATGAGTTACTTTACAGATTCAAGGTAGCTAACTTTTCCAATATGGAT 360  
DB 3189 CTGAGTGTGGGGATGAGTGTCTTCTCAGTTTAGGTTGGCAATTTTCAATATATGAT 3248

QY 361 GAAGATGACATTGAATTGGAACACCAACAAAATCTAAGAAACTGGGAAGAAATCAATTCCA 420  
DB 3249 GAAGATGACATTGAATTGGAACCTGAAAGAAATTCAAAGAACTGGGAGGAGATCAATCCA 3308

QY 421 GAAGTTCACTGGCGAGGAATAGAGGNGNGAAGACAAAAGAACTTGAAGAAATATAT 480  
DB 3309 GAAGAGCAAGAGCGGCGACTAGAAGAGGAGGAGAGACAAAAGAACTGGAGGAAATTTAT 3368

QY 481 ATGCTTCCAAGATGAGAACTGTGCAAAACAGATCAGCTTTAATGGAATCAAGGGAGA 540  
DB 3369 ATGCTCCCGAGATGAGAACTGTGCAAGAGCAGATAAGTTTCAATGGAAGTGAAGGGAG 3428

QY 541 TGCAGTAGGAGCAGAGATATTTCTGGATCTGATGTATTCATTCCTCAGAAAAGAAAACA 600  
DB 3429 CGGAGTAGAAGCAGGAGATATTTCTGGATCTGATGTATTCATCTCGGAAAGAAACGG 3488

QY 601 CCAAAAAACGTTGGAGCAGCAACTATTCCTCCCTGAAAACATTAAGGATTTAGTGAT 660

DB 3489 CCGAAGAAACGTGGGCGACCCCGCACTATCCCTCGGGAGATAATTAAGGATTTAGTGAT 3548

QY 661 CGAGAGATTAGACGATTTATCAAGAGTTTCAAGAAATTTGGTGGCCCACTTGAAGAGTTA 720

DB 3549 GCGAGATTTCGGGGTTTATCAAGAGTTTATAGAAATTTGGTGGCCCTCGGAGAGTTA 3608

QY 721 GATGCTATAGCTAGAGATGCTGAGCTAGTGTATTAATCTGAAACAGACCTTAGACCTCTG 780

DB 3609 GATGCAATTTGCTCGAGATGCTGAATTTGGTTGTAAGTCAAGAAACAGATCTCAGAACAT 3668

QY 781 GGAGAACTTTGTCATATAANTGATGCTAATTAAGGCTTTAAATGATAATGACTTTGGTCAAGA 840

DB 3669 GGAGAACTGCTGCATATAATGTTGTTAAAGCTTTTAAAGAGACAGTTCTTCAGGAACAGAG 3728

QY 841 AGAACAGGTGGTAGATTTTCGGAAGAGTTTAAAGGCCCAACATTCGGAATACAGAGGTGCAG 900

DB 3729 CGAGAGGTGGCAGAGCTTGGAAAGTGAAGGGGCCAACATTCGCGATCTCTGGAGTCCAA 3788

QY 901 GTGAATGCAAGCTAGTCTATTTCTCACGAAGAAGAGTTGGCACCATTTGCATTAATCGATT 960

DB 3789 GTGAATGCCAAGCTGCTCATTTGCCCATGAGGATGAGCTGATCCCTCTGCATAGTCCATC 3848

QY 961 CTTTCAGATCCAGAAAGAAAGAGATATGTCATCCCATACACACACCAAGAGCAGTCTAT 1020

DB 3849 CCTTCGGACCGGAGGAGAGGAGCATATACAATCCCTCGCATACAAAGGCTGCGCAC 3908

QY 1021 TTTGATATAGATTGGGTAAAGAGATGATTCATCTCTTATAGGCATCTATGATAT 1080

DB 3909 TTTGATATAGATTGGGGCAAGAGATGATTTAAATTTGTTTAAATGTTCTATGATGAT 3968

QY 1081 GGTATGGCAGTTGGGAAATGATAAAATGGATCTCTGATCTCAGTTTGACACAGAGATT 1140

DB 3969 GGCTATGGAAGCTGGGAATGATTAATTTGGATCCAGACCTCAGTTTAACACACAGATT 4028

QY 1141 TTACTGTGATGATCCAGATAAGAAACCCAGGCTTAAGCAGTTACAGACTCGTGCGAGATTAC 1200

DB 4029 CTTCCAGATGATCCTGATATAAAACCCAGGCAAGCAAGTACAGACCGCTGCGAGACTAC 4088

QY 1201 CTCATTAATTAAGTAAAGACCTTGCAGAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1260

DB 4089 CTCATTAATTAAGTAAAGACCTTGCAGAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 4148

QY 1261 GGCAATTTCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1313

DB 4149 GGAGGTTCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 4201

RESULT 12  
FAY12942

LOCUS FAY12942 615 bp DNA linear VRT 07-FEB-2002  
DEFINITION Ficedula albicollis CHDIW gene, exon.  
ACCESSION Y12942  
VERSION Y12942.1 GI:2765289  
KEYWORDS CHDIW gene; chromodomain-helicase-DNA-binding protein.  
SOURCE Ficedula albicollis  
ORGANISM Ficedula albicollis

REFERENCE 1 (sites)  
AUTHORS Ellegren, H. and Fridolfsson, A.K.  
TITLE Male-driven evolution of DNA sequences in birds  
JOURNAL Nat. Genet. 17 (2), 182-184 (1997)  
MEDLINE 97467729

REFERENCE 2 (bases 1 to 615)  
AUTHORS Fridolfsson, A.K.  
TITLE Direct Submission  
JOURNAL Submitted (21-APR-1997) A.K. Fridolfsson, Swedish Univ. of  
Agricltural Sciences, Animal Breeding and Genetics, Box 597, S-751  
24, Uppsala, SWEDEN

FEATURES  
source Location/Qualifiers  
1..615

7/240  
8.26  
11

ECV12940		LOCUS	Emberiza citrinella CHD1W gene, exon.	615 bp	DNA	linear	VRT 07-FEB-2002
DEFINITION		Y12940	Emberiza citrinella CHD1W gene, exon.				
ACCESSION		Y12940	Emberiza citrinella CHD1W gene, exon.				
VERSION		Y12940.1	GI:2765287				
KEYWORDS		CHD1W gene; chromodomain-helicase-DNA-binding protein.					
SOURCE		yellowhammer.					
ORGANISM		Emberiza citrinella					
REFERENCE		1 (sites)	Emberiza citrinella				
AUTHORS		Ellegren, H. and Fridolfsson, A.K.					
TITLE		Male-driven evolution of DNA sequences in birds					
JOURNAL		Nat. Genet. 17 (2), 182-184 (1997)					
MEDLINE		97467729					
REFERENCE		2 (bases 1 to 615)	Fridolfsson, A.K.				
AUTHORS		Direct Submission					
TITLE		Submitted (21-APR-1997) A.K. Fridolfsson, Swedish Univ. of					
JOURNAL		Agricultural Sciences, Animal Breeding and Genetics, Box 597, S-751					
BASE COUNT		239 a 94 c 152 g 130 t					
ORIGIN							
FEATURES		source					
1..615		/organism="Emberiza citrinella"					
/db_xref="taxon:37595"							
/chromosome="W"							
/tissue_type="blood"							
/dev_stage="adult"							
1..615		/gene="CHD1W"					
1..615		/gene="CHD1W"					
<1..>615		/gene="CHD1W"					
/codon_start=1							
/product="chromodomain-helicase-DNA-binding protein"							
/db_xref="GI:18642397"							
/translation="FEIDILERAKKKWLDHLVIOIMDTCKTVLHGTSPSSTPPNK							
BELSAILKFGABELFKPEGEPEPEQEMDIDELIKRAETRENEPGPLVGDLELLSKFK							
VANFSNMDEDDIELEFPQNLRWEEIIPVORRLEEREROKELEIYMLPRMNCCK							
QISPNSEGGKCSRRLRYSGSDSDISERKPKRKRPTIPRENI"							
BASE COUNT		239 a 94 c 152 g 130 t					
ORIGIN							
Query Match		41.8%; Score 549.8; DB 5; Length 615;					
Best Local Similarity		93.0%; Pred. No. 2.le-113;					
Matches 572; Conservative		0; Mismatches 43; Indels 0; Gaps 0;					
QY		31 GAAGAAGATATCTTGAAGAGCCAGAAAGAGATGGTGTAGATCAATTTAGTATTGATTCAG	90				
DB		1 GAAGAAGATATCTTGAAGAGCCAGAAAGAGATGGTGTAGATCAATTTAGTATTGATTCAG	60				
QY		91 AGAATGGACACACAGGGAAGACTGTACTACATACAGGCTCTACTCTTCAAGCTCAACA	150				
DB		61 AGAATGGACACTACAGGGAAGACTGTACTACATACAGGCTCTACTCTTCAAGCTCAACA	120				
QY		151 CCTTTTAATGAAGAGAGTTATCAGCAATTTTGAAGTTGGTGGTGGAGAACTTTTAA	210				
DB		121 CCTTTTAATGAAGAGAGTTATCAGCAATTTTGAAGTTGGTGGTGGAGAACTTTTAA	180				
QY		211 GAAGCTGAANNNGAAGAGAGAGCTCTAGGAGATGGATATAGATGAATCTCTGAAGAG	270				
DB		181 GAAGCTGAAGGGAAGAGAGAGAGCTCTAGGAGATGGATATAGATGAATCTCTGAAGAG	240				
QY		271 NCTGAAACTCGAAGAAATGAGTTCAGGCCCTTAAGTCTAGGAGATGAGTCTTTCACAG	330				
DB		241 CCGTGAAGAAATGAGTTCAGGCCCTTAAGTCTAGGAGATGAGTCTTTCACAG	300				
QY		331 TTCAGGTAGCTAACTTTTCCCAATATGATGAAGATGACATTTGAATGGAACGACAA	390				
DB		301 TTCAGGTAGCTAACTTTTCCCAATATGATGAAGATGACATTTGAATGGAACGACAA	360				
QY		391 AATCTAAGAACTGGGAAGAAATCATTTCCAGAACTTCAGTGGCGACGAATAGAGGGGNG	450				
DB		361 AATTTAAGAACTGGGAAGAAATCATTTCCAGAACTTCAGTGGCGACGAATAGAGAAAG	420				
QY		451 GAAAGCAAAAGAACTTTGAAGAAATATATATGCTTCCAGAACTGAGAACTGTGCAAAA	510				
DB		421 GAAAGCAAAAGAACTTTGAAGAAATATATATGCTTCCAGAACTGAGAACTGTGCAAAA	480				
QY		511 CAGATCAGCTTTTAAATGAAGATGAGGAGATGAGTGGAGGAGAGATATCTTGGATCT	570				
DB		481 CAGATCAGCTTTTAAATGAAGATGAGGAGATGAGTGGAGGAGAGATATCTTGGATCT	540				
QY		571 GATAGTATCTCATCTCAGAAAGAAACGACCAAAACGTTGGACGACCAAGCAATATT	630				
DB		541 GATAGTATCTCATCTCAGAAAGAAACGACCAAAACGTTGGACGACCAAGCAATATT	600				
QY		631 CCCCGTGAAGAACTTT 645					
DB		601 CCTCGAAGAAATATT 615					
RESULT		13					

Db	301	TTCAAGTGGCTAACTTTTCAAATATGGATGAAGATGACATTGAGTTGAAACCCAGGACAA	360	Matches	558;	Conservative	0;	Mismatches	57;	Indels	0;	Gaps	0;
QY	391	AATCTAGAAACTGGGAAGAAATCATTTCCAGAAAGTTTCAGTGGCGAGCAATAGAGGGNNG	450	QY	31	GAAGAAATATTTCTTCAAAGAGCCCAAGAAAGATGGTGTAGATCATTTAGTATTTCAG	90						
Db	361	AATTAGAAACTGGGAAGAAATCATTTCCAGAAAGTTTCAGTGGCGAGCAATAGAGAGGAA	420	Db	1	GAAGAAATATTTCTTGAACGAGCCCAAGAAAGATGGTGTAGATCATTTAGTATTTCAG	60						
QY	451	GAAGACAAAAAGAACTTGAAGAAATATATATGCTTCCAAAGATGAGAAACTGTGCACAAA	510	QY	91	AGAATGGACACACACAGGGGAAAACTGTACTACATACAGGCTCTACTCCTTCAAGCTCAACA	150						
Db	421	GAAGACAAAAAGAACTTGAAGAAATATATATGCTTCCAAAGATGAGAAACTGTGCACAAA	480	Db	61	AGAATGGACACTACAGGGGAAAACTGTGTACATACAGGCTCTACTCCTTCAAGCTCAACA	120						
QY	511	CAGATCAGCTTTAATGGAATGAAGGAGATGCAGTAGGACGACGAGATATCTGGATCT	570	QY	151	CCTTTTAAATAAGAGAGAGTTATCAGCAATTTTGAAGTTTGGTGTCTGAGGAACTTTTAAA	210						
Db	481	CAGATCAACTTTAATGGAAGGAGAAACCCAGTAGGAGAGATATCTGGATCT	540	Db	121	CCTTTTAAATAAGGAGAGTTGTCTGCAATTTTGAAGTTTGGTGTCTGAGGAACTTTTAAA	180						
QY	571	GATAGTATTCCATCTCAGAAAGAAACGACCAAAACGTTGAGGACGACGAGAACTATT	630	QY	211	GAACCTGAANNNGAAGAGAGGAGCTCAGGAGATGGATATAGATGAATCCTGAAGAGG	270						
Db	541	GATAGGACTCCATCTCAGAAAGAAACGACCAAAACGTTGAGGACGACGAGAACTATT	600	Db	181	GAACGGAAGGGAAGAGCAGGAGGCTCAAGAGATGGATATAGATGAATCTGAAGAGG	240						
QY	631	CCCCGTGAAAAACATT	645	QY	271	NCTGAAACTCGAGAAATAGTTCAGGCCCATTAACCTGTAGGAGATGAGTTACTTTTCACAG	330						
Db	601	CCTCGAGAAAAATATT	615	Db	241	GCTGAAACTCGGAAAAATGAGCCAGGTCATTAACCTGTAGGAGATGAGTTCTTTCGCAG	300						
RESULT	14			QY	331	TTCAAGGTAGCTAACTTTTCCAAATATGGATGAAGATGACATTGAAATGGAACACGAGACAA	390						
LOCUS	FAY12939	Ficedula albicollis	615 bp	DNA	linear	VRT 07-FEB-2002							
DEFINITION	Ficedula albicollis	CHD1Z gene, exon.		Db	301	TTCAAGGTGGCCCAATCTTCCAAATATGGATGAAGATGATTAATTTGGAACCCAGGAAAGA	360						
ACCESSION	Y12939			QY	391	AATCTAAGAAACTGGGAAGAAATCATTTCCAGAAAGTTTCAGTGGCGAGCAATAGAGGGNNG	450						
VERSION	Y12939.1	GI:2765286		Db	361	AATCTAAGAAACTGGGAAGAAATCATTTCCAGAAAGTTTCAGTGGCGAGCAATAGAGAGGAG	420						
KEYWORDS	CHD1Z gene; chromodomain-helicase-DNA-binding protein.			QY	451	GAAGACAAAAAGAACTTGAAGAAATATATATGCTTCCAAAGATGAGAAACTGTGCACAAA	510						
SOURCE	Ficedula albicollis.			Db	421	GAAGACAAAAAGAACTTGAAGAAATATATATGCTTCCAAAGATGAGAAACTGTGCACAAA	480						
ORGANISM	Ficedula albicollis			QY	511	CAGATCAGCTTTAATGGAATGAAGGAGATGCAGTAGGACGACGAGATATCTTGGATCT	570						
REFERENCE	1 (sites)			Db	481	CAGATTAGCTTCAATGGCAGTGAGGGGAGGCGCAGTAGGATAGAAATATCTTGGATCT	540						
AUTHORS	Ellegren,H. and Fridolfsson,A.K.			QY	571	GATAGTATTCCATCTCAGAAAGAAACGACCAAAACGTTGAGGACGACGAGAACTATT	630						
TITLE	Male-driven evolution of DNA sequences in birds			Db	541	GATAGTACTCCATCTCAGAAAGAAACGACCAAAACGTTGAGGACGACGAGAACTATT	600						
JOURNAL	Nat. Genet. 17 (2), 182-184 (1997)			QY	631	CCCCGTGAAAAACATT	645						
MEDLINE	97467729			Db	601	CCTCGAGAAAAATATT	615						
REFERENCE	2 (bases 1 to 615)			RESULT	15								
AUTHORS	Fridolfsson,A.K.			ECY12941									
TITLE	Direct Submission			LOCUS	Emberiza citrinella	615 bp	DNA	linear	VRT 07-FEB-2002				
JOURNAL	Submitted (21-APR-1997) A.K. Fridolfsson, Swedish Univ. of Agricultural Sciences, Animal Breeding and Genetics, Box 597, S-751 24, Uppsala, SWEDEN			DEFINITION	Emberiza citrinella	CHD1Z gene, exon.							
FEATURES	Location/Qualifiers			ACCESSION	Y12941								
source	1. .615			VERSION	Y12941.1	GI:2765288							
	/organism="Ficedula albicollis"			KEYWORDS	CHD1Z gene; chromodomain-helicase-DNA-binding protein.								
	/db_xref="taxon:59894"			SOURCE	yellowhammer.								
	/chromosome="z"			ORGANISM	Emberiza citrinella								
	/tissue_type="blood"			REFERENCE	1 (sites)								
	/dev_stage="adult"			AUTHORS	Ellegren,H. and Fridolfsson,A.K.								
exon	1. .615			TITLE	Male-driven evolution of DNA sequences in birds								
gene	/gene="CHD1Z"			JOURNAL	Nat. Genet. 17 (2), 182-184 (1997)								
CDS													

Job time: 26872 sec

/db\_xref="taxon:37595"  
/chromosome="2"  
/tissue\_type="blood"  
/dev\_stage="adult"  
1..615  
/gene="CHD12"  
1..615  
/gene="CHD12"  
1..615  
/gene="CHD12"  
1..615  
/codon\_start=1  
/product="chromodomain-helicase-DNA-binding protein"  
/protein\_id="CAD23052.1"  
/db\_xref="GI:18642396"  
/translation="EEDILERAKKMLVHLVQIMDDTGTGKVLHTGSTPSSSTPFFK  
EELSAILKFGAELEFKPEGEQPOEMDIDEILKRAETRENEPGPLTVGDELISOFK  
VANFSNMDEDDIELEPERNSRNWEEIIPVORRIIEEERKELEIYMLPRMNCAC  
QISFNGSEGRSRSRYSGSDSDSISERKRPKRGRPTIPRENT"  
BASE COUNT 232 a 95 c 157 g 131 t  
ORIGIN

Query Match 39.5%; Score 519.4; DB 5; Length 615;  
Best Local Similarity 89.9%; Pred. No. 1.4e-106;  
Matches 553; Conservative 0; Mismatches 62; Indels 0; Gaps 0;  
QY 31 GAAGAGATATTCCTGAAGAGCCCAAGAAAAGATGGTGTAGATCATTTAGTGATTCAG 90  
Db 1 GAAGAGATATTCCTGAAGAGCCCAAGAAAAGATGGTGTAGATCATTTAGTGATTCAG 60  
QY 91 AGAATGGACACACAGGAGAAACTGTACTACATACAGGCTCTACTCTTCAAGCTCAACA 150  
Db 61 AGAATGGACACTACAGGAGAAACTGTGTACATACAGGCTCTACTCTTCAAGCTCAACA 120  
QY 151 CCTTTTAATAGGAGAGTTATCAGCAATTTTGAAGTTTGGTCTGAGGAACCTTTTAAA 210  
Db 121 CCTTTTAACAGAGAGAGTTGTGCGCAATTTTAAAGTTTGGTCTGAGGAACCTTTTAAA 180  
QY 211 GAACCTGAANNNGAAGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 270  
Db 181 GAACCTGAAGGGGAAGAACAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 240  
QY 271 NCTGAACACTCGAGAAATGAGTCAGGCCCATTAACCTGAGGAGATGAGTTACTTCCACAG 330  
Db 241 GCCGAACCTCGGAAATGAGCCAGTCCATTAACCTGAGGAGATGAGTTCTTTCGCAG 300  
QY 331 TTCAGGTAGCTTAACCTTCCATATGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 390  
Db 301 TTCAGGTGCGCACTTCTCCAATATGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 360  
QY 391 AATCTAAGAACTGGAGAAATCATTCAGAGTTTCAGTGGCGAGCAATAGAGGGGNG 450  
Db 361 AATCAAGAAATGGAGAAATCATTCAGAGTTTCAGAGTCCAGAGCGAAGAAAGAGAGAG 420  
QY 451 GAAGACAAAAGAACTGGAAGAAATATATATGCTTCCAGAAATGAGAAACTGTGCAAAA 510  
Db 421 GAAGACAAAAGAACTGGAAGAAATATATATGCTTCCAGAAATGAGAAACTGTGCAAAA 480  
QY 511 CAGATCAGCTTTAATGGAAATGAAGGAGATGCAGTAGGAGCAGAGATATTTCTGGATCT 570  
Db 481 CAGATTAGCTTTCAGAGCAAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 540  
QY 571 GATAGTGATTCATCTCAGAAAGAAACAGCAAAAAACGTGGAGCAGCAGCACTATT 630  
Db 541 GATAGTGATTCATCTCAGAAAGAAAGGCAAAAAACGTGGAGCAGCAGCACTATT 600  
QY 631 CCCGTGAACACT 645  
Db 601 CCTCGAGAAATATT 615

Search completed: August 2, 2002, 21:00:31



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.  
OM nucleic - nucleic search, using sw model  
Run on: August 2, 2002, 20:30:20 ; Search time 12674 Seconds  
(without alignments)  
252.625 Million cell updates/sec  
Title: US-08-973-363-4  
Perfect score: 153  
Sequence: 1 ATTTTACCTGATGATCAGA.....CACAGAGACTTGCTGTGCA 153  
Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues  
Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl.\*  
1: gb\_ba.\*  
2: gb\_htg.\*  
3: gb\_in.\*  
4: gb\_om.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_sts.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vi.\*  
15: em\_ba.\*  
16: em\_fun.\*  
17: em\_hum.\*  
18: em\_in.\*  
19: em\_mu.\*  
20: em\_om.\*  
21: em\_or.\*  
22: em\_ov.\*  
23: em\_pat.\*  
24: em\_ph.\*  
25: em\_pl.\*  
26: em\_ro.\*  
27: em\_sts.\*  
28: em\_un.\*  
29: em\_vi.\*  
30: em\_htg\_hum.\*  
31: em\_htg\_inv.\*  
32: em\_htg\_other.\*  
33: em\_htgo\_inv.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No. Query Score Match Length DB ID Description

1	153	100.0	153	6	A58685	A58685 Sequence 4
2	140.2	91.6	153	6	A58684	A58684 Sequence 3
3	133.8	87.5	153	6	A58686	A58686 Sequence 5
4	105.6	69.0	153	6	A58683	A58683 Sequence 2
5	100.4	65.6	1316	6	A58696	A58696 Sequence 15
6	95	62.1	6608	6	A58691	A58691 Sequence 10
7	95	62.1	6872	5	AF004397	AF004397 Gallus ga
8	69.8	45.6	2292	5	DL4316	DL4316 delta-cryst
9	69	45.1	5349	10	MUSCHD1X	L10410 Mouse DNA-b
10	61.2	40.0	5947	9	AF006513	AF006513 Homo sapi
C 11	61	39.9	101220	9	AC092372	AC092372 Homo sapi
12	61	39.9	134365	9	AC012624	AC012624 Homo sapi
13	61	39.9	143079	9	AC021449	AC021449 Homo sapi
14	61	39.9	145659	2	AC008531	AC008531 Homo sapi
15	61	39.9	193446	2	AC091946	AC091946 Homo sapi
C 16	61	39.9	219258	9	AC022121	AC022121 Homo sapi
C 17	61	39.9	276181	2	AC092382	AC092382 Homo sapi
18	47.8	31.2	348	5	AF060702	AF060702 Struthio
19	41.2	26.9	1311	6	AR029026	AR029026 Sequence
20	39.6	25.9	446	6	AX333515	AX333515 Sequence
21	39.6	25.9	7764	9	AF006514	AF006514 Homo sapi
C 22	38.8	25.9	6893	2	AC087641	AC087641 Homo sapi
C 23	39.6	25.9	177951	2	AC074383	AC074383 Homo sapi
C 24	39.6	25.9	183658	9	AC013394	AC013394 Homo sapi
25	35	22.9	6242	3	DROCHD1A	L77907 Drosophila
C 26	35	22.9	14530	2	AC019885	AC019885 Drosophila
27	35	22.9	163403	3	AC007765	AC007765 Drosophila
28	35	22.9	164920	3	AC092187	AC092187 Drosophila
C 29	35	22.9	188002	9	AC008069	AC008069 Homo sapi
30	35	22.9	230277	2	AL645930	AL645930 Mus muscu
C 31	35	22.9	306436	3	AE003581	AE003581 Drosophila
C 32	34.2	22.4	84406	2	AC094996	AC094996 Rattus no
C 33	34.2	22.4	96746	2	AC094598	AC094598 Rattus no
34	34	22.2	104848	2	AC004588	AC004588 Homo sapi
35	34	22.2	159849	2	AC021454	AC021454 Homo sapi
C 36	34	22.2	160169	2	AC051664	AC051664 Homo sapi
C 37	34	22.2	163915	2	AC087451	AC087451 Homo sapi
C 38	34	22.2	166867	9	AP003733	AP003733 Homo sapi
C 39	33.6	22.0	60894	2	AC090930	AC090930 Homo sapi
C 40	33.4	21.8	44114	2	AC101790	AC101790 Mus muscu
C 41	33.4	21.8	104577	2	AC101790	AC101790 Mus muscu
C 42	33.4	21.8	139744	9	AL133513	AL133513 Rattus no
C 43	33.4	21.8	176524	9	AL133513	AL133513 Human DNA
C 44	33.4	21.8	179303	2	AC010726	AC010726 Homo sapi
C 45	33.4	21.8	191415	2	AL627302	AL627302 Mus muscu

ALIGNMENTS

RESULT	1	A58685	Sequence 4 from Patent WO9639505.	153 bp	DNA	linear	PAT 06-MAR-1998
LOCUS		A58685					
DEFINITION		A58685					
ACCESSION		A58685					
VERSION		A58685.1	GI:3714248				
KEYWORDS		unidentified.					
SOURCE		unclassified.					
ORGANISM		unclassified.					
REFERENCE		1 (bases 1 to 153)					
AUTHORS		Griffiths R. and Tiwari B.					
TITLE		AVIAN GHD GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN BIRDS					
JOURNAL		Patent: WO 9639505-A 4 12-DEC-1996;					
COMMENT		ISIS INNOVATION (GB)					
FEATURES		Other publication AU 5906996 961224.					
source		Location/Qualifiers					
		1..153					
		/organism="unidentified"					
		/db_xref="taxon:32644"					
BASE COUNT		56 a	36 c	31 g	30 t		
ORIGIN							

TITLE	AVIAN GHD GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN BIRDS									
JOURNAL	Patent: WO 9639505-A 5 12-DEC-1996; ISIS INNOVATION (GB)									
COMMENT	Other publication AU 5906996 961224.									
FEATURES	Location/Qualifiers source 1..153 /organism="unidentified" /db_xref="taxon:32644" 28 t									
BASE COUNT	60 a	34 c	31 g	28 t						
ORIGIN										
Query Match	87.5%; Score 133.8; DB 6; Length 153;									
Best Local Similarity	92.3%; Pred. No. 6.3e-30;									
Matches	141; Conservative	0; Mismatches	12; Indels	0; Gaps						
Qy	1	ATTTTACTGATGATCCAGATAGAAACCCGAGCTAAGCAGTTCAGACGACCAAGAAACCC	60							
Db	1	ATTTTACTGATGATCCAGATAGAAACCCGAGCTAAGCAGTTCAGACGACCAAGAAACCA	60							
Qy	61	CAGCTAAGCAGTTACAGACCCCGTGCAGATTACCTCATTAATTACTGATAAAGACCTT	120							
Db	61	CAGCAAGCAGTTGTCAGACCCCGTGCAGATTACCTCATTAATTACTGATAAAGACCTT	120							
Qy	121	GCAAGAAAGGACACAGAGACTTGTGCTGGTGCA	153							
Db	121	GCAAGAAAGGACACAGAGACTTGTGCTGGTGCA	153							
RESULT 4										
A58683	A58683 153 bp DNA linear PAT 06-MAR-1998									
LOCUS	Sequence 2 from Patent WO9639505.									
DEFINITION	A58683									
ACCESSION	A58683.1 GI:371246									
VERSION	unidentified.									
KEYWORDS	unclassified.									
SOURCE	1 (bases 1 to 153) Griffiths,R. and Tiwari,B. AVIAN GHD GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN BIRDS									
ORGANISM	unclassified.									
REFERENCE	1 (bases 1 to 153) Griffiths,R. and Tiwari,B. AVIAN GHD GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN BIRDS									
AUTHORS	Patent: WO 9639505-A 2 12-DEC-1996;									
TITLE	ISIS INNOVATION (GB)									
JOURNAL	Other publication AU 5906996 961224.									
COMMENT	Location/Qualifiers									
FEATURES	source 1..153 /organism="unidentified" /db_xref="taxon:32644" 28 t									
BASE COUNT	61 a	37 c	27 g	28 t						
ORIGIN										
Query Match	69.0%; Score 105.6; DB 6; Length 153;									
Best Local Similarity	80.9%; Pred. No. 2.1e-21;									
Matches	123; Conservative	0; Mismatches	29; Indels	0; Gaps						
Qy	1	ATTTTACTGATGATCCAGATAGAAACCCGAGCTAAGCAGTTCAGACGACCAAGAAACCC	60							
Db	1	ATTTTACTGATGATCCAGATAGAAACCCGAGCTAAGCAGTTCAGACGACCAAGAAACCA	60							
Qy	61	CAGCTAAGCAGTTACAGACCCCGTGCAGATTACCTCATTAATTACTGATAAAGACCTT	120							
Db	61	CAAGCAAAACAGTTACAGACCCCGTGCAGATTACCTCATTAATTACTGATAAAGACCTT	120							
Qy	121	GCAAGAAAGGACACAGAGACTTGTGCTGGTGCC	152							
Db	121	GCAAGAAAGGACACAGAGACTTGTGCTGGTGCC	152							
RESULT 5										



[illegible]



DAIARDAELVDKSETDLRLGELVHNGCVKALKOSSGTPRAGRGKVGKPTFRISG  
VQYAKLVSEIMEDIPJLHKISIPSPERKQYTPCHTAAHEDIMWCKSDSNLIG  
VYQYLGWSEIMEDIPJLHKLDPDPKQPAKQLQTRADYLKLLSLAKREA  
QRLGAGGSRKTRAKSKASKIVKEEIKSDSPSEKSEDDDKLNDKSPK  
DRSKSVADPAHVITAGEBPVIAEESDELDQKTSICKERNRPVKAAKQDRPEK  
GUSERELHTROCLIGDITELKESNPEQIKQRKNLWIFVSKTFEDAKHL  
KLYKIAIKKKQSQNSQNSVATVIRINPDMERKENTHDDSSRDSYSDRHLH  
QYHHDHQRHQSGYKKSDSKRQYSGFNGKDRHWDHYRQDSRYSDREKRLDD  
HRSREHRSLEGLDKDRHRSRHSRHHSDHRSSEHTHHKSRDYRYLSDWQDL  
HRAASGSPRSPDQSRQYSGVSPENSAHRSTPHTWSSRKT"  
1739 a 1067 c 1319 q 1224 t

Query Match	45.1%	Score 69	DB 10	Length 5349
Best Local Similarity	80.2%	Pred. No. 1.7e-10		
Matches 81	Conservative	Mismatches 20	Indels 0	Gaps 0
52	AGAAACCCCGAGCTAAGCAGTTACAGACCCGTCGAGATTACCTCATTTAAATTAAGTAAT	111		
4047	AAAAACACCAAGCAAAACAGTTACAGACCCGTCGAGATTACCTATCAAACTACTTAGC	4106		
112	AAAGACCTTGCAGAAAGGAGACAGAGACTTGTGGTGC	152		
4107	AGAGATCTTGCAGAAAGAGAGGCTCAGAGACTTTGTGGTGC	4147		

RESULT	10
AF006513	
LOCUS	5947 bp mRNA linear PRI 27-NOV-1997
DEFINITION	Homo sapiens CHD1 mRNA, complete cds.
ACCESSION	AF006513
VERSION	AF006513.1 GI:2645428
KEYWORDS	
SOURCE	human.
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini, Homnidae; Homo. 1 (bases 1 to 5947) Woodege,T., Basrai,M.A., Barexvanis,A.D., Hietter,P. and Collins,F.S. Characterization of the CHD family of proteins Proc. Natl. Acad. Sci. U.S.A. 94 (21), 11472-11477 (1997)
AUTHORS	
TITLE	
MEDLINE	97470991
REFERENCE	2 (bases 1 to 5947) Woodege,T. Direct Submission Submitted (03-JUN-1997) Laboratory of Gene Transfer, National Human Genome Research Institute, National Institutes of Health, 49 Convent Drive, Bethesda, MD 20892-4442, USA
JOURNAL	

[illegible][illegible]

	Query Match	Best Local Similarity	40.0%;	Score 61.2;	DB 9;	Length 5947;
	Matches	75;	Conservative	0;	Mismatches 23;	Indels 0;
QY	52	AGAAAACCCCGGCTAAGCAGTTACAGACCCGCTGCAGATTACCTCATTAATTAATTCGAAT	111			
DB	4046	AAAAAACCAACGAACAAAGCTTGCAGACCCGCTGCAGACTACCTCATCAAAATTAATTACTAGT	4105			
QY	112	AAAGACCTTGCAGAAAGGAACACAGAGACTTTGCTGG	149			
DB	4106	AGAGATCTTGCAGAAAAGAAAGCTCTTCTGTGGTGGGG	4143			

RESULT	11
AC092372/c	
LOCUS	101220 bp DNA linear PRI 07-DEC-2001
DEFINITION	Homo sapiens chromosome 5 clone RP11-58M12, complete sequence.
ACCESSION	AC092372
VERSION	AC092372.3 GI:17402768
KEYWORDS	HTG.
SOURCE	human.
ORGANISM	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1 (bases 1 to 101220)
AUTHORS	DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE	Direct Submission
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 101220)
AUTHORS	DOE Joint Genome Institute.
TITLE	Direct Submission
JOURNAL	Submitted (03-JUL-2001) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE	3 (bases 1 to 101220)
AUTHORS	DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE	Direct Submission
JOURNAL	Submitted (07-DEC-2001) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT	On Dec 7, 2001 this sequence version replaced gi:15290448.

www.shgc.stanford.edu  
Quality: Phrap Quality >=40 100% of Sequence;  
Estimated Total Number of Errors is 0.  
NOTE: This insert is not the entire sequence of the clone (entire  
sequence is 146,7kb). It is clipped at the overlap with AC012624.  
The number of bases overlapped is 90404.  
Location/Qualifiers:



SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 145659)  
AUTHORS DOE Joint Genome Institute.  
TITLE Sequencing of Human Chromosome 5  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 145659)  
AUTHORS DOE Joint Genome Institute.  
TITLE Direct Submission  
JOURNAL Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
COMMENT On Feb 14, 2001 this sequence version replaced gi:7528342.  
-----Genome Center  
Center: Joint Genome Institute  
Web site: <http://www.jgi.doe.gov>

```
web site: http://www.jg1.doe.gov
-----
```

Project Information  
Center Project Name: 369535  
Center Clone name: CIT-HSPC 480B11

Center clone name: CIT-HSPC\_480B11  
-----  
Summary Statistics  
Consensus quality: 139128 bases at least Q40

Consensus quality: 139128 bases at least Q40  
Consensus quality: 142556 bases at least Q30  
Consensus quality: 143744 bases at least Q20  
Estimated insert size: 149000, culled field gel estimation

Estimated insert size: 148000; pulse field gel estimation  
Estimated insert size: 145059; sum-of-contigs estimation  
Quality coverage: 6.27 in Q20 bases; pulse field gel estimation  
Quality coverage: 5.4 in Q20 bases; sum-of-contigs estimation

Quality coverage: 6.4 in Q20 bases; sum-of-contigs estimation.  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 7 contigs. Gaps between the contigs

\* are represented as runs of  $N$ . The order of the pieces  
 \* is believed to be correct as given, however the sizes  
 \* of the gaps between them are based on estimates that have

- \* provided by the submittor.
- \* This sequence will be replaced
- \* by the finished sequence as soon as it is available and

\* the accession number will be preserved.  
 \* 1 56174: contig of 56174 bp in length  
 \* 56175 56274: gap of unknown length

\* 56275 100874: contig of 44600 bp in length  
\* 100875 100974: gap of unknown length  
\* 100975 113127: contig of 12153 bp in length

\* 113128 113227: gap of unknown length  
\* 113228 118190: contig of 4963 bp in length  
\* 118191 118290: gap of unknown length

\* 118291 119694: contig of 1404 bp in length  
\* 110695 110704: gap of unknown length

\* 123298 123397: gap of unknown length  
\* 123398 145659: contig of 2262 bp in length.

Location/Qualifiers  
1. .145659

```
I. .143039
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTC-480B11"
```

```

42561 a 26309 c 27580 g 48609 t 600 others
/clone_lib="CalTech human BAC library C"
/clone_cdc="00001"

```

[illegible]

h 39.9%; Score 61; DB 2; Length 145659;  
Similarity 82.4%; Pred. NO. 3.5e-08;  
70; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

[illegible]



Mon Aug 5 11:51:52 2002

us-08-973-363-4.rge

Page 9

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.  
OM nucleic - nucleic search, using sw model  
Run on: August 2, 2002, 20:31:05 ; Search time 12674 Seconds  
(without alignments)  
252.625 Million cell updates/sec  
Title: US-08-973-363-5  
Perfect score: 153  
Sequence: 1 ATTTTACCTGATGACCCAGA.....TGCAGAGACTTACTGGTGCA 153  
Scoring table: IDENTIFY\_NUC  
Gapex 10.0 , Gapext 1.0  
Searched: 1797656 seqs, 10463268293 residues  
Total number of hits satisfying chosen parameters: 3595312  
Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries  
Database : GenBank:\*

1: gb\_ba:\*

2: gb\_htg:\*

3: gb\_in:\*

4: gb\_om:\*

5: gb\_ov:\*

6: gb\_pat:\*

7: gb\_ph:\*

8: gb\_pl:\*

9: gb\_pr:\*

10: gb\_ro:\*

11: gb\_sts:\*

12: gb\_sy:\*

13: gb\_un:\*

14: gb\_vi:\*

15: em\_ba:\*

16: em\_fun:\*

17: em\_hum:\*

18: em\_in:\*

19: em\_mu:\*

20: em\_om:\*

21: em\_or:\*

22: em\_ov:\*

23: em\_pat:\*

24: em\_ph:\*

25: em\_pl:\*

26: em\_ro:\*

27: em\_sts:\*

28: em\_un:\*

29: em\_vi:\*

30: em\_htg\_hum:\*

31: em\_htg\_inv:\*

32: em\_htg\_other:\*

33: em\_htgo\_inv:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	ID	Description
------------	-------	-------	-------	--------	----	-------------

1	153	100.0	153	6	A58686	A58686 Sequence 5
2	133.8	87.5	153	6	A58685	A58685 Sequence 4
3	130.6	85.4	153	6	A58684	A58684 Sequence 3
4	105.6	69.0	153	6	A58683	A58683 Sequence 2
5	88.6	57.9	6608	6	A58691	A58691 Sequence 10
6	88.6	57.9	6872	5	AF004397	AF004397 Gallus ga
7	87.6	57.3	1316	6	A58696	A58696 Sequence 15
8	69	45.1	5349	10	MUSCHD1X	L10410 Mouse DNA-b
9	68.2	43.6	2292	5	D14316	D14316 delta-cryst
10	66	43.1	5947	9	AC006513	AC006513 Homo sapi
11	66	43.1	101220	9	AC092372	AC092372 Homo sapi
12	66	43.1	134365	9	AC012624	AC012624 Homo sapi
13	66	43.1	143079	2	AC021449	AC021449 Homo sapi
14	66	43.1	145659	2	AC008531	AC008531 Homo sapi
15	66	43.1	193446	2	AC009146	AC009146 Homo sapi
16	66	43.1	219258	9	AC022121	AC022121 Homo sapi
17	66	43.1	276181	2	AC092382	AC092382 Homo sapi
18	43	28.1	348	5	AF060702	AF060702 Struthio
19	42.8	28.0	1311	6	AR029026	AR029026 Sequence
20	37.8	24.7	208531	2	AC087560	AC087560 Mus muscu
21	35.2	23.0	164275	9	AC005157	AC005157 Homo sapi
22	35	22.9	6242	3	DROCHD1A	L77907 Drosophila
23	35	22.9	14530	2	AC019885	AC019885 Drosophila
24	35	22.9	163403	3	AC007765	AC007765 Drosophila
25	35	22.9	164920	3	AC092187	AC092187 Drosophila
26	35	22.9	306436	3	AE003581	AE003581 Drosophila
27	34.8	22.7	6029	3	AB015484	AB015484 Dugesia j
28	34.8	22.7	179303	2	AC016087	AC016087 Homo sapi
29	34.6	22.6	169660	9	AP003547	AP003547 Homo sapi
30	34.2	22.4	446	6	AX333515	AX333515 Sequence
31	34.2	22.4	7764	9	AF006514	AF006514 Homo sapi
32	34.2	22.4	19514	8	SPAC39E6	266525 S. pombe chr
33	34.2	22.4	36315	8	SPAC30E6	AL136538 S. pombe c
34	34.2	22.4	69893	2	AC087641	AC087641 Homo sapi
35	34.2	22.4	177951	2	AC074383	AC074383 Homo sapi
36	34.2	22.4	183658	9	AC013394	AC013394 Homo sapi
37	34.2	22.4	207705	2	AL590630	AL590630 Mus muscu
38	34	22.2	207705	2	AL590630	AF220294 Mus muscu
39	33.8	22.1	175759	10	AF220294	AC097516 Homo sapi
40	33.6	22.0	159064	9	AC097516	AC021356 Homo sapi
41	33.6	22.0	163072	30	AC021356	AC021356 Homo sapi
42	33.4	21.8	67068	2	AC017801	AC017801 Drosophila
43	33.4	21.8	69551	2	AC101956	AC101956 Mus muscu
44	33.4	21.8	69551	2	AC101956	AC101956 Mus muscu
45	33.4	21.8	171226	3	AC008144	AC008144 Drosophila

ALIGNMENTS

RESULT	1	A58686	Sequence 5 from Patent WO9639505.	153 bp	DNA	linear	PAT 06-MAR-1998
LOCUS	A58686						
DEFINITION	A58686						
ACCESSION	A58686						
VERSION	A58686.1	GI:3714249					
KEYWORDS	unidentified.						
SOURCE	unclassified.						
ORGANISM	1 (bases 1 to 153)						
REFERENCE	Griffiths R. and Tiwari, B.						
AUTHORS	AVIAN GHD GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN						
TITLE	BIRDS						
JOURNAL	Patent: WO 9639505-A 5 12-DEC-1996;						
COMMENT	ISIS INNOVATION (GB) 5906996 961224.						
FEATURES	Other Publication AU 5906996 961224.						
source	Location/Qualifiers						
	1..153						
	/organism="unidentified"						
	/db_xref="taxon:32644"						
BASE COUNT	60 a	34 c	31 g	28 t			
ORIGIN							

```
Query Match      100.0%; Score 153; DB 6; Length 153;
Best Local Similarity 100.0%; Pred. No. 1.1e-34;
Matches 153; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATTTTACCTGATGACCCAGATAGAAACACAGGCAAGCAGTTGCAGACCAAGAAACCA 60
|||||
Db 1 ATTTTACCTGATGACCCAGATAGAAACACAGGCAAGCAGTTGCAGACCAAGAAACCA 60
|||||

QY 61 CAGGCAAAACAGTTGCAGACCCGTCAGATTACCTCATTAAATTAATTAAGACCTT 120
|||||
Db 61 CAGGCAAAACAGTTGCAGACCCGTCAGATTACCTCATTAAATTAATTAAGACCTT 120
|||||

QY 121 GCAAGAAAAGAGTGCAAGACTTACTGTGCA 153
|||||
Db 121 GCAAGAAAAGAGTGCAAGACTTACTGTGCA 153
|||||

RESULT 2
LOCUS A58685 153 bp DNA linear PAT 06-MAR-1998
DEFINITION Sequence 4 from Patent WO9639505.
ACCESSION A58685
VERSION A58685.1 GI:3714248
KEYWORDS
SOURCE
ORGANISM
unidentified.
unclassified.
1 (bases 1 to 153)
Griffiths,R. and Tiwari,B.
AVIAN GHD GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN
BIRDS
JOURNAL
Patent: WO 9639505-A 4 12-DEC-1996;
ISIS INNOVATION (GB)
Other publication AU 5906996 961224.
COMMENT
FEATURES
source
/organism="unidentified"
/db_xref="taxon:32644"

BASE COUNT 56 a 36 c 31 g 30 t
ORIGIN

Query Match      87.5%; Score 133.8; DB 6; Length 153;
Best Local Similarity 92.2%; Pred. No. 4.7e-29;
Matches 141; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 1 ATTTTACCTGATGACCCAGATAGAAACACAGGCAAGCAGTTGCAGACCAAGAAACCA 60
|||||
Db 1 ATTTTACCTGATGATCCAGATAGAAACCCAGGCTAGCAGTTACAGACCAAGAAACCC 60
|||||

QY 61 CAGGCAAAACAGTTGCAGACCCGTCAGATTACCTCATTAAATTAATTAAGACCTT 120
|||||
Db 61 CAGGCTAAGCAGTTACAGACCCGTCAGATTACCTCATTAAATTAATTAAGACCTT 120
|||||

QY 121 GCAAGAAAAGAGTGCAAGACTTACTGTGCA 153
|||||
Db 121 GCAAGAAAAGAGCAGACAGACTTGTGTGCA 153
|||||

RESULT 3
LOCUS A58684 153 bp DNA linear PAT 06-MAR-1998
DEFINITION Sequence 3 from Patent WO9639505.
ACCESSION A58684
VERSION A58684.1 GI:3714247
KEYWORDS
SOURCE
ORGANISM
unidentified.
unclassified.
1 (bases 1 to 153)
Griffiths,R. and Tiwari,B.
AVIAN GHD GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN
BIRDS
JOURNAL
Patent: WO 9639505-A 2 12-DEC-1996;
ISIS INNOVATION (GB)
Other publication AU 5906996 961224.
COMMENT
FEATURES
source
/organism="unidentified"
/db_xref="taxon:32644"

BASE COUNT 61 a 37 c 27 g 28 t
ORIGIN

Query Match      69.0%; Score 105.6; DB 6; Length 153;
Best Local Similarity 80.9%; Pred. No. 9.5e-21;
Matches 123; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 1 ATTTTACCTGATGACCCAGATAGAAACACAGGCAAGCAGTTGCAGACCAAGAAACCA 60
|||||
Db 1 ATTTCTCCAGATGATCCTGTATATAAACCCACAGCAAGCAACAGTACAGACCAAAAAACCA 60
|||||

QY 61 CAGGCAAAACAGTTGCAGACCCGTCAGATTACCTCATTAAATTAATTAAGACCTT 120
|||||
Db 61 CAAGCAAAACAGTTACAGACCCGTCAGACTACCTCATTAAATTAATTAAGACCTT 120
|||||

QY 121 GCAAGAAAAGAGTGCAAGACTTACTGTGTC 152
|||||
Db 121 GCAAGAAAAGAGGCTCAGAGACTTTGTGTGTC 152
|||||

RESULT 5
```





/translation="MELKKCNHCYLKPPDDNEFYKQALQHLIRSGKLILDLK  
LILRERGRVLIIFSQMRMLDILAEYLKYPFQRLDGSIGKELRQALDHFNAEG  
SEDFCLSTRAGGLGINLASADTVIFDSWNPQNDLQAARAHRIQKQOVNIYRL  
VTGVSVEEDILERAKKMWLDHLVQIORMDTGKTLVHTGSTPSSSTPFKEELSLALK  
FGAEELFPEGEQEQEPMDEILKRAETRENEPGLTVGDELLSQFVAFNEMD  
EDIELEPERNSRWEIIPESORRRIEERKELEIYMLPMRNCARQISFNGSE  
GRSRRYSGSDSDSITERRKPKRGRPTIPRENIKGSDAETRRFISYKKGCGP  
LERLDVARDLAEVSHBELPLHKSIPSDPERKQVIPCPTKAAHFDIMGKEDDSN  
LLVGIYVAYKWSMOCKVNGSRSQLNTEILPDDPRTPRNSVRPVQTTSLNY"  
BASE COUNT 827 a 385 c 546 g 534 t  
ORIGIN

Query Match 44.6%; Score 68.2; DB 5; Length 2292;  
Best Local Similarity 85.4%; Pred. No. 9.8e-10;  
Matches 88; Conservative 0; Mismatches 13; Indels 2; Gaps 1;

Qy 53 AGAACCCAGGCAACGAGTTCGACCGCGTACCTATTAAATTAAGTAATA 112  
Db 1883 AGAATCCAGCAACGAGTTCGACCGCGTACCTATTAAATTAAGTAATA 1942  
Qy 113 AGACCTTGCAGAAAGAGT--GCAAGACCTACTGTGCA 153  
Db 1943 AGACCTTGCAGAAAGAGTTCGACCGCGTACCTATTAAATTAAGTAATA 1985

RESULT 10  
AF006513  
LOCUS Homo sapiens CHD1 mRNA, complete cds.  
DEFINITION AF006513  
ACCESSION AF006513.1 GI:2645428  
VERSION  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 5947)  
AUTHORS Woodage, T., Bearai, M.A., Bayevanis, A.D., Hietter, P. and Collins, F.S.  
TITLE Characterization of the CHD family of proteins  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 94 (21), 11472-11477 (1997)  
MEDLINE 97470991  
REFERENCE 2 (bases 1 to 5947)  
AUTHORS Woodage, T.  
TITLE Direct Submission  
JOURNAL Submitted (03-JUN-1997) Laboratory of Gene Transfer, National Human Genome Research Institute, National Institutes of Health, 49 Convent Drive, Bethesda, MD 20892-4442, USA

FEATURES  
source 1. .5947  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="5"  
/map="5q15-21; near WI-5811"  
1. .5947  
/gene="CHD1"  
164. .5293  
/codon\_start=1  
/gene="CHD1"  
/product="CHD1"  
/protein\_id="AA087381.1"  
/db\_xref="GI:2645429"

translation="MNGHSDSEVRNNSGSSQSDSDSGSGSGSSGSSGSSGSS  
SQSGSDSDSGSGSSQSDSESDTRENKVOAKPKVDCAPKWKSSPILAVORSAIL  
KQOQOQOQOQOASNSSESDSSSEVPKPKHDEDMQSGSGSSQSS  
GDSSEDERKSSCDESDYVKNKPKQNSKSKKTLGKQRIQDASER  
DDDEYDNDKSSRQAVNYSYKEDPMKTDSDLEVEDVQPPPEERFIER  
FMDRIGRGKATGATTVIYVADGDNDNAGEKKEPGEIQYLKNGSHHINWET  
EMTLKQNVGMKLDNYKKQETKRLNANSPEDYINCCQELIDDLHKOIQLV  
RIANSGKSAAGTPIYCKWGLFYSECWEDGALISKRFACIDITFRNSKQITVG  
FKCKVLRPRFRFAPKQSTIGGHELELDYQLNGLNWLNWISWCKNSCILADEM  
GLGKTIQITSLNLFHFHOLYGPFLIVVPLUSTLTWSQREIQTWASQMNNAVYLGDN

gene  
CDS  
164. .5293  
/codon\_start=1  
/gene="CHD1"  
/product="CHD1"  
/protein\_id="AA087381.1"  
/db\_xref="GI:2645429"

translation="MNGHSDSEVRNNSGSSQSDSDSGSGSGSSGSSGSSGSS  
SQSGSDSDSGSGSSQSDSESDTRENKVOAKPKVDCAPKWKSSPILAVORSAIL  
KQOQOQOQOQOASNSSESDSSSEVPKPKHDEDMQSGSGSSQSS  
GDSSEDERKSSCDESDYVKNKPKQNSKSKKTLGKQRIQDASER  
DDDEYDNDKSSRQAVNYSYKEDPMKTDSDLEVEDVQPPPEERFIER  
FMDRIGRGKATGATTVIYVADGDNDNAGEKKEPGEIQYLKNGSHHINWET  
EMTLKQNVGMKLDNYKKQETKRLNANSPEDYINCCQELIDDLHKOIQLV  
RIANSGKSAAGTPIYCKWGLFYSECWEDGALISKRFACIDITFRNSKQITVG  
FKCKVLRPRFRFAPKQSTIGGHELELDYQLNGLNWLNWISWCKNSCILADEM  
GLGKTIQITSLNLFHFHOLYGPFLIVVPLUSTLTWSQREIQTWASQMNNAVYLGDN

SRMIRTHWTHOTKRLKFNILLTYELLKDKAFILGNLWAFIVGDEAHLKNDSD  
LLYKTLIDKSNHLLITGTPLQNSLKEWLLHFMPEKFSWEDFEEHKGREY  
YASLHLELFPDLIRRVKDKSLPAKVEILKRMESALQKQYKWLITRNYKALSG  
SKSTSGFINIMELKKCNHCYLKPPDDNEFYKQALQHLIRSGKLILDLKLI  
DLRERGRVLIIFSQMRMLDILAEYLKYPFQRLDGSIGKELRQALDHFNAEG  
RFRERGRVLIIFSQMRMLDILAEYLKYPFQRLDGSIGKELRQALDHFNAEG  
DFCLSTRAGGLGINLASADTVIFDSWNPQNDLQAARAHRIQKQOVNIYRLV  
KGSVEEDILERAKKMWLDHLVQIORMDTGKTLVHTGSTPSSSTPFKEELSLALK  
EDIELEPERNSRWEIIPESORRRIEERKELEIYMLPMRNCARQISFNGSE  
GRSRRYSGSDSDSITERRKPKRGRPTIPRENIKGSDAETRRFISYKKGCGP  
LERLDVARDLAEVSHBELPLHKSIPSDPERKQVIPCPTKAAHFDIMGKEDDSN  
LLVGIYVAYKWSMOCKVNGSRSQLNTEILPDDPRTPRNSVRPVQTTSLNY"  
BASE COUNT 2130 a 1004 c 1243 g 1570 t  
ORIGIN

Query Match 43.1%; Score 66; DB 9; Length 5947;  
Best Local Similarity 87.8%; Pred. No. 4.4e-09;  
Matches 72; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 52 AAGAACCCAGGCAACGAGTTCGACCGCGTACCTATTAAATTAAGTAATA 111  
Db 4046 AAGAACCCAGGCAACGAGTTCGACCGCGTACCTATTAAATTAAGTAATA 4105  
Qy 112 AAGACCTTGCAGAAAGAGAGAG 133  
Db 4106 AGAGATCTTGCAGAAAGAGAGAG 4127

RESULT 11  
AC092372/c  
LOCUS Homo sapiens chromosome 5 clone RP11-58M12, complete sequence.  
DEFINITION AC092372  
ACCESSION AC092372.3 GI:17402768  
VERSION  
KEYWORDS HTG.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 101220)  
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.  
TITLE Direct Submission  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 101220)  
AUTHORS DOE Joint Genome Institute.  
TITLE Direct Submission  
JOURNAL Submitted (03-JUL-2001) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

REFERENCE 3 (bases 1 to 101220)  
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.  
TITLE Direct Submission  
JOURNAL Submitted (07-DEC-2001) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

COMMENT On Dec 7, 2001 this sequence version replaced gi:15290448.  
Draft Sequence Produced by DOE Joint Genome Institute  
www.jgi.doe.gov  
www.hgc.stanford.edu  
Finishing Completed at Stanford Human Genome Center  
Quality: Phrap Quality >=40 100% of Sequence;  
Estimated Total Number of Errors is 0.  
NOTE: This insert is not the entire sequence of the clone (entire  
sequence is 146.7Kb). It is clipped at the overlap with AC012624.  
The number of bases overlapped is 90404.  
Location/Qualifiers

FEATURES

Db 118328 AGAGATCTTCACAAAAGAG 118349 .

RESULT 13	
AC0211449	
LOCUS	AC0211449
DEFINITION	143079 bp DNA linear HTG 10-SEP-2000
ACCESSION	AC0211449
VERSION	GI:10047806
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	1 (bases 1 to 143079)
TITLE	Birren, B., Linton, L., Nusbaum, C. and Lander, E.
REFERENCE	Homo sapiens, clone RP11-58M12
AUTHORS	Unpublished
	2 (bases 1 to 143079)
	Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Beckerly, R., Beda, F., Boguslavskiy, L., Boukhalter, B., Brown, A., Burkett, G., Castle, A., Choepli, J., Colangelo, M., Collins, S., Collymore, A., Cooke, P., DeArelano, K., Dewar, K., Domino, M., Doyle, M., Fenestor, J., Ferreira, P., FitzHugh, W., Forrest, C., Gage, D., Galagan, J., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Landers, T., Lehoczy, J., Levine, R., Liu, C., Locke, K., MacDonald, P., Marguis, N., McEwan, P., McGurk, A., McKernan, K., McPheeters, R., Meldrim, J., Meneus, L., Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., Oliver, T. M., Peterson, K., Pierre, N., Pisani, C., Pollara, V., Raymond, C., Riley, R., Rothman, D., Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.
TITLE	Direct Submission
JOURNAL	Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT	On Sep 10, 2000 this sequence version replaced gi:7407963. All repeats were identified using RepeatMasker: Smit, A.F.A. & Green, P. (1996-1997) <a href="http://ftp.genome.washington.edu/RM/RepeatMasker.html">http://ftp.genome.washington.edu/RM/RepeatMasker.html</a> ----- Genome Center Center: Whitehead Institute/ MIT Center for Genome Research Center code: WIBR Web site: <a href="http://www-seq.wi.mit.edu">http://www-seq.wi.mit.edu</a> Contact: <a href="mailto:sequence_submissions@genome.wi.mit.edu">sequence_submissions@genome.wi.mit.edu</a> ----- Project Information Center project name: L5154 Center clone name: 58_M_12 ----- Summary Statistics Sequencing vector: M13; M77815; 100% of reads Chemistry: Dye-terminator Big Dye; 100% of reads Assembly program: Phrap; version 0.960731 Consensus quality: 134743 bases at least Q40 Consensus quality: 139227 bases at least Q30 Consensus quality: 140814 bases at least Q20 Insert size: 144000; agarose-fp Insert size: 142179; sum-of-contigs Quality coverage: 4.6 in Q20 bases; agarose-fp Quality coverage: 4.7 in Q20 bases; sum-of-contigs ----- * NOTE: This is a 'working draft' sequence. It currently * consists of 10 contigs. The true order of the pieces * is not known and their order in this sequence record is * arbitrary. Gaps between the contigs are represented as * runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence * as soon as it is available and the accession number will * be preserved. * * 1 38820: contig of 38820 bp in length



```
Db 46851 AGAGATCTTGCAAAAAGAG 46872
|||||
RESULT 15
AC091946 193446 bp DNA linear HTG 09-JUN-2001
LOCUS Homo sapiens chromosome 5 clone RP11-360I2, *** SEQUENCING IN
DEFINITION PROGRESS ***, 33 unordered pieces.
ACCESSION AC091946
VERSION AC091946.1 GI:14333882
KEYWORDS HTG; HTGS_PRAISE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 193446)
DOE Joint Genome Institute.
Sequencing of Human Chromosome 5
Unpublished
2 (bases 1 to 193446)
DOE Joint Genome Institute.
Direct Submission
Submitted (09-JUN-2001) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 544799
Center clone name: RPCI-11_360I2
-----
Summary Statistics
Consensus quality: 157767 bases at least Q40
Consensus quality: 180259 bases at least Q30
Consensus quality: 184175 bases at least Q20
Estimated insert size: 204590; agarose-fp estimation
Estimated insert size: 190246; sum-of-contigs estimation
Quality coverage: 5.38 in Q20 bases; agarose-fp estimation
Quality coverage: 5.79 in Q20 bases; sum-of-contigs estimation.
NOTE: This is a 'working draft' sequence. It currently
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1358: contig of 1358 bp in length
1359 1458: gap of unknown length
1459 2667: contig of 1209 bp in length
2668 2767: gap of unknown length
2768 3823: contig of 1056 bp in length
3824 3923: gap of unknown length
3924 5920: contig of 1997 bp in length
5921 6020: gap of unknown length
6021 7494: contig of 1474 bp in length
7495 7594: gap of unknown length
7595 9096: contig of 1502 bp in length
9097 9196: gap of unknown length
9197 10260: contig of 1064 bp in length
10261 10360: gap of unknown length
10361 12460: contig of 2100 bp in length
12461 12560: gap of unknown length
12561 14611: contig of 2051 bp in length
14612 14711: gap of unknown length
14712 16381: contig of 1670 bp in length
16382 16481: gap of unknown length
16482 17968: contig of 1487 bp in length
17969 18068: gap of unknown length
18069 20434: contig of 2366 bp in length
```

```
* 20435 20534: gap of unknown length
* 20535 23515: contig of 2981 bp in length
* 23516 23615: gap of unknown length
* 23616 27563: contig of 3948 bp in length
* 27564 27663: gap of unknown length
* 27664 30987: contig of 3324 bp in length
* 30988 31087: gap of unknown length
* 31088 36065: contig of 4978 bp in length
* 36066 36165: gap of unknown length
* 36166 40978: contig of 4813 bp in length
* 40979 41078: gap of unknown length
* 41079 45663: contig of 4585 bp in length
* 45664 51745: contig of 5982 bp in length
* 51746 51845: gap of unknown length
* 51846 57359: contig of 5514 bp in length
* 57360 57459: gap of unknown length
* 57460 67881: contig of 10422 bp in length
* 67882 74132: contig of 6151 bp in length
* 74133 74232: gap of unknown length
* 74233 79795: contig of 5563 bp in length
* 79796 79895: gap of unknown length
* 79896 87511: contig of 7616 bp in length
* 87512 87611: gap of unknown length
* 87612 92791: contig of 5180 bp in length
* 92792 92792: gap of unknown length
* 92793 102794: contig of 9903 bp in length
* 102795 102894: gap of unknown length
* 102895 110866: contig of 7972 bp in length
* 110867 110966: gap of unknown length
* 110967 117571: contig of 6605 bp in length
* 117572 117671: gap of unknown length
* 117672 123738: contig of 6067 bp in length
* 123739 123838: gap of unknown length
* 123839 130583: contig of 6745 bp in length
* 130584 130683: gap of unknown length
* 130684 141544: contig of 10861 bp in length
* 141545 141644: gap of unknown length
* 141645 169109: contig of 27465 bp in length
* 169110 169209: gap of unknown length
* 169210 193446: contig of 24237 bp in length.
FEATURES
    source
    1..193446
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /chromosome="5"
    /clone="RP11-360I2"
    /clone_lib="RPCI human BAC library 11"
BASE COUNT 57571 a 35252 c 37061 g 60360 t 3202 others
ORIGIN
```

```
Query Match 43.1%; Score 66; DB 2; Length 193446;
Best Local Similarity 87.8%; Pred. No. 4.5e-09;
Matches 72; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
Qy 52 AAGAAACCAAGGCAACAGTCGACACCGTCGAGATTACCTCATTAATTAATTAAT 111
Db 87679 AAAAAACCAAGCAACAGTCGACACCGTCGAGATTACCTCATTAATTAATTAAT 111
Qy 112 AAAGACCTTGCAGAAAAAGAG 133
Db 87739 AGAGATCTTGCAAAAAGAG 87760
```

Search completed: August 2, 2002, 20:31:51  
Job time: 25152 sec



us-08-973-363-5.rge

Mon Aug 5 11:51:54 2002

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 2, 2002, 14:23:50 ; Search time 10310.5 seconds  
(without alignments)  
1467.418 Million cell updates/sec

Title: US-08-973-363-1

Perfect score: 723

Sequence: 1 CCCGTCGAGGTTTCAAGG.....CTTGAAACCTCCGACCGGG 723

Scoring table:

CGCGC\_NUC  
Gapop 60.0, Gapext 60.0

Searched: 1797656 seqs, 10463268293 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl.\*

1: gb.ba.\*

2: gb.bhg.\*

3: gb.in.\*

4: gb.ov.\*

5: gb.om.\*

6: gb.pat.\*

7: gb.ph.\*

8: gb.pl.\*

9: gb.pr.\*

10: gb.ro.\*

11: gb.sts.\*

12: gb.sy.\*

13: gb.un.\*

14: gb.vi.\*

15: em.ba.\*

16: em.fun.\*

17: em.hum.\*

18: em.in.\*

19: em.mu.\*

20: em.om.\*

21: em.or.\*

22: em.ov.\*

23: em.pat.\*

24: em.ph.\*

25: em.pl.\*

26: em.ro.\*

27: em.sts.\*

28: em.un.\*

29: em.vi.\*

30: em.htg\_hum.\*

31: em.htg\_inv.\*

32: em.htg\_other.\*

33: em.htgo\_inv.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
------------	-------	-------	--------	----	-------------

1	723	100.0	723	6	A58682
2	32	4.4	239130	2	AC079420
3	31	4.3	100325	2	AC103438
4	31	4.3	158755	2	AC091404
5	31	4.3	161092	2	AC100725
6	31	4.3	162071	2	AC096932
7	31	4.3	169467	2	AC096932
8	31	4.3	170536	9	HS518K17
9	31	4.3	173591	9	AC022101
10	31	4.3	182328	9	AP001646
11	31	4.3	199311	2	AL662788
12	31	4.3	240682	2	AC093465
13	30	4.1	65400	2	AC101852
14	30	4.1	67166	2	AC102249
15	30	4.1	67166	2	AC102249
16	30	4.1	88013	9	AC090511
17	30	4.1	96627	9	AC073284
18	30	4.1	101728	9	HS109088
19	30	4.1	131957	2	AL645986
20	30	4.1	143811	2	AC013617
21	30	4.1	152144	2	AC016293
22	30	4.1	154280	9	AC025561
23	30	4.1	157134	2	AC106642
24	30	4.1	158408	9	AL391821
25	30	4.1	160367	9	AC009542
26	30	4.1	165951	2	AC104677
27	30	4.1	169758	2	AL356139
28	30	4.1	170271	2	AC097587
29	30	4.1	172038	2	AC096926
30	30	4.1	174768	9	AC009330
31	30	4.1	174800	9	AC009137
32	30	4.1	178424	9	AC069020
33	30	4.1	180763	2	AC067826
34	30	4.1	180763	2	AC067826
35	30	4.1	180821	2	AC016791
36	30	4.1	181539	2	AL104132
37	30	4.1	182427	9	AL162390
38	30	4.1	182931	9	AC100856
39	30	4.1	196024	2	AC018338
40	30	4.1	198719	2	AC015716
41	30	4.1	208580	2	AC080078
42	30	4.1	228516	2	AC008594
43	30	4.1	242487	2	AL645938
44	29	4.0	697	10	CPANP
45	29	4.0	16015	9	AF059650

ALIGNMENTS

RESULT	1	A58682	Sequence 1 from Patent WO9639505.	723 bp	DNA	linear	PAT 06-MAR-1998
LOCUS	A58682	Sequence 1 from Patent WO9639505.					
DEFINITION	A58682	Sequence 1 from Patent WO9639505.					
ACCESSION	A58682	Sequence 1 from Patent WO9639505.					
VERSION	A58682.1	GI:3714245					
KEYWORDS		unidentified.					
SOURCE		unclassified.					
ORGANISM		unclassified.					
REFERENCE		1 (bases 1 to 723)					
AUTHORS		Griffiths, R. and Tiwari, B.					
TITLE		AVIAN GHD GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN BIRDS					
JOURNAL		Patent: WO 9639505-A 1 12-DEC-1996;					
COMMENT		ISIS INNOVATION (GB)					
FEATURES		Other publication AU 5906996 961224.					
		Location/Qualifiers					
		1..723					
		/organism="unidentified"					
		/db_xref="taxon:32644"					
		188 a 109 c 164 g 262 t					
		BASE COUNT					
		ORIGIN					





Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K.,  
Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N.,  
Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R.,  
Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C.,  
Watlington, S., Williams, G., Williamson, A., Wlezyk, R., Wooden, S.,  
Worley, K., Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zorrilla, S., Nelson, D.,  
Weinstock, G. and Gibbs, R.

TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

## COMMENT

Unpublished  
2 (bases 1 to 100325)  
Worley, K.C.  
Direct Submission  
Submitted (25-NOV-2001) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Dec 20, 2001 this sequence version replaced gi:17064227.  
----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: GNR  
Center clone name: CH230-160E23  
----- Summary Statistics  
Assembly program: Phrap; version 0.990329First call to  
findPhrapList  
Consensus quality: 82281 bases at least Q40  
Consensus quality: 88976 bases at least Q30  
Consensus quality: 94579 bases at least Q20  
Estimated insert size: 78830; sum-of-contigs estimation  
Quality coverage: 0x in Q20 bases; agarose-fp estimation  
Quality coverage: 0.9x in Q20 bases; sum-of-contigs estimation  
-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank/draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 47 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 4899: contig of 4899 bp in length  
\* 4900 4999: gap of unknown length  
\* 5000 9265: contig of 4266 bp in length  
\* 9266 9365: gap of unknown length  
\* 9366 14917: contig of 5551 bp in length  
\* 14917 15016: gap of unknown length  
\* 15017 19118: contig of 4102 bp in length  
\* 19119 19218: gap of unknown length  
\* 19219 22751: contig of 3533 bp in length  
\* 22752 22851: gap of unknown length  
\* 22852 25552: contig of 2801 bp in length  
\* 25553 25752: gap of unknown length  
\* 25753 28904: contig of 3152 bp in length  
\* 28905 29004: gap of unknown length  
\* 29005 32083: contig of 3079 bp in length  
\* 32084 32183: gap of unknown length  
\* 32184 35571: contig of 3388 bp in length  
\* 35572 35671: gap of unknown length  
\* 35672 37220: contig of 1549 bp in length  
\* 37221 37321: gap of unknown length  
\* 37321 39031: contig of 1711 bp in length  
\* 39032 39131: gap of unknown length  
\* 39132 41551: contig of 2420 bp in length  
\* 41552 41651: gap of unknown length  
\* 41652 43298: contig of 1647 bp in length  
\* 43299 43398: gap of unknown length  
\* 43399 45054: contig of 1656 bp in length  
\* 45055 45154: gap of unknown length  
\* 45155 46224: contig of 1070 bp in length

46324: gap of unknown length  
46325 48386: contig of 2052 bp in length  
48387 48486: gap of unknown length  
48487 50138: contig of 1652 bp in length  
50139 50238: gap of unknown length  
50239 52581: contig of 2343 bp in length  
52582 52681: gap of unknown length  
52682 55707: contig of 3026 bp in length  
55708 55807: gap of unknown length  
55808 57227: contig of 1420 bp in length  
57228 57327: gap of unknown length  
57328 58484: contig of 1157 bp in length  
58485 58584: gap of unknown length  
58585 60039: contig of 1455 bp in length  
60040 60139: gap of unknown length  
60140 61159: contig of 1020 bp in length  
61160 61259: gap of unknown length  
61260 63385: contig of 2126 bp in length  
63386 63485: gap of unknown length  
63486 64660: contig of 1175 bp in length  
64661 64760: gap of unknown length  
64761 66498: contig of 1738 bp in length  
66499 66598: gap of unknown length  
66599 67895: contig of 1297 bp in length  
67896 67996: gap of unknown length  
67997 70321: contig of 2326 bp in length  
70322 70421: gap of unknown length  
70422 71783: contig of 1362 bp in length  
71784 71883: gap of unknown length  
71884 72999: contig of 1116 bp in length  
72999 73000: gap of unknown length  
73000 74541: contig of 1442 bp in length  
74542 74641: gap of unknown length  
74642 76364: contig of 1723 bp in length  
76365 76464: gap of unknown length  
76465 78499: contig of 2035 bp in length  
78499 78500: gap of unknown length  
78500 80253: contig of 1654 bp in length  
80254 80353: gap of unknown length  
80354 82294: contig of 1941 bp in length  
82295 82394: gap of unknown length  
82395 84444: contig of 2050 bp in length  
84445 84544: gap of unknown length  
84545 86158: contig of 1613 bp in length  
86159 86257: gap of unknown length  
86258 87683: contig of 1426 bp in length  
87684 87783: gap of unknown length  
87784 88890: contig of 1107 bp in length  
88891 88990: gap of unknown length  
88991 90673: contig of 1683 bp in length  
90674 90773: gap of unknown length  
90774 92235: contig of 1462 bp in length  
92236 92335: gap of unknown length  
92336 93522: contig of 1187 bp in length  
93523 93622: gap of unknown length  
93623 94779: contig of 1157 bp in length  
94780 94879: gap of unknown length  
94880 96090: contig of 1211 bp in length  
96091 96190: gap of unknown length  
96191 97230: contig of 1040 bp in length  
97231 97330: gap of unknown length  
97331 98892: contig of 1562 bp in length  
98893 98993: gap of unknown length  
98994 100325: contig of 1333 bp in length.

FEATURES  
Source

1..100325  
/organism="Rattus norvegicus"  
/db\_xref="taxon:10116"  
/clone="CH230-160E23"

BASE COUNT 27941 a 19883 c 20639 g 27225 t 4637 others  
ORIGIN

```

* 32724 55171: contig of 22448 bp in length
* 55172 55271: gap of unknown length
* 55272 86367: contig of 31096 bp in length
* 86368 86467: gap of unknown length
* 86468 158755: contig of 72288 bp in length.
FEATURES
      Location/Qualifiers
          1..158755
             /organism="Sus scrofa"
             /db_xref="taxon:9823"
             /clone="RP44-74011"
             /clone_lib="RP44"
          1..3649
             /note="assembly_fragment"
             clone_end:SP6
             vector_side:left"
          3750..13662
             /note="assembly_fragment"
          13763..32623
             /note="assembly_fragment"
          32724..55171
             /note="assembly_fragment"
             clone_end:T7
             vector_side:left"
          55272..86367
             /note="assembly_fragment"
          86468..158755
             /note="assembly_fragment"
BASE COUNT 41826 a 38951 c 37930 g 39519 t 529 others
ORIGIN
Query Match 4.3%; Score 31; DB 2; Length 158755;
Best Local Similarity 100.0%; Pred. No. 7.5e-06;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 445 ACCTCTTTTTTTTTTTTTTTTTTTTGGCTT 475
|||||TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT

Db 32502 ACCTCTTTTTTTTTTTTTTTTTTTTGGCTT 32532

RESULT 5
AC100725/c
LOCUS
DEFINITION Mus musculus clone RP24-192F16, WORKING DRAFT SEQUENCE, 13 linear
FEATURES
      Location/Qualifiers
          161092 bp
          DNA
          HTG 22-NOV-2001
          unorderd pieces.
          AC100725
          AC100725.1 GI:17048092
          HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
          house mouse.
          Mus musculus
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
          1 (bases 1 to 161092)
          Birren,B., Linton,L., Nusbaum,C. and Lander,E.
          Mus musculus, clone RP24-192F16
          Unpublished
          2 (bases 1 to 161092)
          Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
          Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhalil,B.,
          Brown,A., Camarata,J., Campotiano,A., Chang,J., Chazaro,B.,
          Choepel,X., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
          Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
          Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gargyala,S.,
          Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
          Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
          Jones,C., Kamat,A., Karatas,A., Kellis,C., LaRocque,K.,
          Lamazares,R., Landers,T., Lehotzky,J., Levine,R., Liu,G.,
          MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
          McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
          Menes,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
          Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
          Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
          Raymond,C., Retta,R., Ribbeck,M., Riley,R., Rise,C., Rogov,P.,

```

Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

# TITLE JOURNAL

Submitted (22-NOV-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www.seq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

----- Project Information

Center project name: L17410

Center clone name: 192\_F\_16

----- Summary Statistics

Sequencing vector: Plasmid; N/A; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 156876 bases at least Q40

Consensus quality: 158247 bases at least Q30

Consensus quality: 158979 bases at least Q20

Insert size: 157000; agarose-fp

Insert size: 159892; sum-of-contigs

Quality coverage: 9.0 in Q20 bases; agarose-fp

Quality coverage: 8.8 in Q20 bases; sum-of-contigs

-----

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 13 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 9907: contig of 9907 bp in length  
\* 9508 10007: gap of 100 bp  
\* 10008 11803: contig of 1796 bp in length  
\* 11804 11903: gap of 100 bp  
\* 11904 14164: contig of 2261 bp in length  
\* 14165 14264: gap of 100 bp  
\* 14265 16540: contig of 2276 bp in length  
\* 16541 16640: gap of 100 bp  
\* 16641 18624: contig of 1984 bp in length  
\* 18625 18724: gap of 100 bp  
\* 18725 22393: contig of 3569 bp in length  
\* 22394 22393: gap of 100 bp  
\* 22394 28253: contig of 5860 bp in length  
\* 28254 28353: gap of 100 bp  
\* 28354 68690: contig of 40337 bp in length  
\* 68691 68790: gap of 100 bp  
\* 68791 88661: contig of 19871 bp in length  
\* 88662 88761: gap of 100 bp  
\* 88762 109433: contig of 20672 bp in length  
\* 109434 109533: gap of 100 bp  
\* 109534 127910: contig of 18377 bp in length  
\* 127911 128010: gap of 100 bp  
\* 128011 149944: contig of 21934 bp in length  
\* 149945 150044: gap of 100 bp  
\* 150045 161092: contig of 11048 bp in length.

## FEATURES source

1. .161092  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="Rp24-192F16"  
/clone\_lib="RPCI-24 Male Mouse BAC"  
misc\_feature  
1. .9907  
/note="assembly\_fragment"

clone\_end:SP6  
vector\_side:left  
10008..11803  
/note="assembly\_fragment"  
11904..14164  
/note="assembly\_fragment"  
14265..16540  
/note="assembly\_fragment"  
16641..18624  
/note="assembly\_fragment"  
18725..22293  
/note="assembly\_fragment"  
22394..28253  
/note="assembly\_fragment"  
28354..68690  
/note="assembly\_fragment"  
68791..88661  
/note="assembly\_fragment"  
88762..109433  
/note="assembly\_fragment"  
109534..127910  
/note="assembly\_fragment"  
128011..149944  
/note="assembly\_fragment"  
150045..161092  
/note="assembly\_fragment"  
clone\_end:r7  
vector\_side:right  
BASE COUNT 44315 a 34609 c 34525 g 46427 t 1216 others  
ORIGIN

Query Match 4.3% Score 31; DB 2: Length 161092;

Best Local Similarity 100.0%; Pred. No. 7.5e-06; Mismatches 0; Indels 0; Gaps 0;

QY 440 AAACCACTCTTTTTTTTTTTTTTTTTTTTTTTT 470

|||||TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT

DB 52256 AAACCACTCTTTTTTTTTTTTTTTTTTTTTTTT 52226

RESULT 6

AC096932/c

LOCUS 162071 bp DNA linear HTG 21-DEC-2001  
DEFINITION Rattus norvegicus chromosome Renin clone CH230-57H13, \*\*\*  
SEQUENCING IN PROGRESS \*\*\*, 66 unordered pieces.

AC096932

AC096932.3 GI:17975665

HTG: HTGS\_PHASE1.

SOURCE Norway rat.

ORGANISM

Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.

REFERENCE

AUTHORS

1 (bases 1 to 162071)  
Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C.,  
Alsbrooks, S.L., Amaral, H.C., Are, J.R., Banks, T., Barbaria, J.,  
Benton, J., Binage, K., Blankenburg, K., Bonnin, D., Bouck, J.,  
Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C.,  
Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carlton, T.F.,  
Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R.,  
Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C.,  
Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C.,  
Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O.,  
Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H.,  
Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C.,  
Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J.,  
Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T.,  
Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S.,  
Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A.,  
Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Holloway, C.,  
Hollins, B., Homsif, F., Howard, S., Huber, J., Hulyk, S., Hume, J.,  
Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jollivet, S.,



Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Loizeg, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapa, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhiney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokkenko, S., Oguh, M., Okunolu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Picken, R., Primus, E., Pu, L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojokan, I., Rolfe, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shoshitani, N., Sisson, T., Sodergren, E., Sonaika, T., Sparks, A., Stanley, H., Tang, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczyk, R., Woodden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstock, G. and Gibbs, R.

TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

2 (bases 1 to 162071)

Worley, K.C.

Submitted (03-OCT-2001)

Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Dec 21, 2001 this sequence version replaced gi:17062391.

COMMENT

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GGKB

Center clone name: CH230-57H13

----- Summary Statistics

Assembly program: Phrap: version 0.990329First call to findPhrapList

Consensus quality: 128926 bases at least Q40

Consensus quality: 136110 bases at least Q30

Consensus quality: 142174 bases at least Q20

Estimated insert size: 129056: sum-of-contigs estimation

Quality coverage: 0x in Q20 bases; agarose-fp estimation

Quality coverage: 2.4x in Q20 bases; sum-of-contigs estimation

-----

NOTE: Estimated insert size may differ from sequence length

(see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).

NOTE: This is a 'working draft' sequence. It currently

consists of 66 contigs. The true order of the pieces

is not known and their order in this sequence record is

arbitrary. Gaps between the contigs are represented as

runs of N, but the exact sizes of the gaps are unknown.

This record will be updated with the finished sequence

as soon as it is available and the accession number will

be preserved.

1 8087: contig of 8087 bp in length

8088 8187: gap of unknown length

8188 16770: contig of 8583 bp in length

16771 16870: gap of unknown length

16871 21956: contig of 5086 bp in length

21957 22058: gap of unknown length

22057 26653: contig of 4596 bp in length

26653 32174: gap of unknown length

32174 32274: contig of 5422 bp in length

32275 35763: gap of unknown length

35764 35864: gap of unknown length

35864 40879: contig of 5016 bp in length

40880 40979: gap of unknown length

40980 45170: contig of 4191 bp in length

45171 45270: gap of unknown length

45271 48431: contig of 3161 bp in length

48432 48531: gap of unknown length

48532 52173: contig of 3642 bp in length

52174 52273: gap of unknown length

52274 55653: contig of 3379 bp in length

55653 55752: gap of unknown length

55753 60083: contig of 4331 bp in length

60084 64504: gap of unknown length

64505 64604: gap of unknown length

64605 67428: contig of 2823 bp in length

67429 70187: contig of 2859 bp in length

70188 70287: gap of unknown length

70288 73313: contig of 3029 bp in length

73314 73416: gap of unknown length

73417 75449: contig of 2034 bp in length

75450 75549: gap of unknown length

75550 77924: contig of 2375 bp in length

77925 78025: gap of unknown length

78026 80940: contig of 2916 bp in length

80941 81041: gap of unknown length

81042 83926: contig of 2886 bp in length

83927 84026: gap of unknown length

84027 86287: contig of 2261 bp in length

86288 86387: gap of unknown length

86388 88052: contig of 1665 bp in length

88053 88152: gap of unknown length

88153 89716: contig of 1564 bp in length

89717 89816: gap of unknown length

89817 91057: contig of 1240 bp in length

91058 91157: gap of unknown length

91158 92732: contig of 1576 bp in length

92733 92832: gap of unknown length

92833 95059: contig of 2227 bp in length

95060 95159: gap of unknown length

95160 97802: contig of 2643 bp in length

97803 97902: gap of unknown length

97903 99786: contig of 1884 bp in length

99787 99886: gap of unknown length

99887 100914: contig of 1028 bp in length

100915 101014: gap of unknown length

101015 102380: contig of 1366 bp in length

102381 102480: gap of unknown length

102481 104222: contig of 1742 bp in length

104223 104322: gap of unknown length

104323 106585: contig of 2263 bp in length

106586 106885: gap of unknown length

106886 109066: contig of 2381 bp in length

109067 109166: gap of unknown length

109167 111958: contig of 2792 bp in length

111959 112058: gap of unknown length

112059 115010: contig of 2952 bp in length

115011 115110: gap of unknown length

115111 116967: contig of 1857 bp in length

116968 117067: gap of unknown length

117069 119078: contig of 2011 bp in length

119079 121701: contig of 2523 bp in length

121702 121801: gap of unknown length

121802 123473: contig of 1672 bp in length

123474 123573: gap of unknown length

123574 124815: contig of 1242 bp in length

124816 126077: gap of unknown length

126078 126178: contig of 1162 bp in length

126179 127825: contig of 1648 bp in length

127826 127925: gap of unknown length

127926 129483: contig of 1558 bp in length

129484 130640: gap of unknown length

130641 130740: gap of unknown length







Consensus quality: 199198 bases at least Q40  
 Consensus quality: 199267 bases at least Q30  
 Consensus quality: 199292 bases at least Q20  
 Insert size: 199311; sum-of-contigs  
 Insert size: 185916; 8.7% error; agarose-fp  
 Quality coverage: 16.09x in Q20 bases; sum-of-contigs Quality  
 coverage: 17.83x in Q20 bases; agarose-fp  
 -----  
 \* NOTE: This is a 'working draft' sequence.  
 \* This sequence will be replaced  
 \* by the finished sequence as soon as it is available and  
 \* the accession number will be preserved.  
 Location/Qualifiers  
 1..199311  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="9"  
 /clone="RP11-778P13"  
 /clone\_lib="RPC1-11.3"  
 1..199311  
 /note="assembly\_fragment:04061"  
 clone\_end:SP6  
 vector\_side:left"  
 BASE COUNT 58307 a 36635 c 39266 g 65103 t  
 ORIGIN

# FEATURES

source

misc\_feature

Query Match 4.3%; Score 31; DB 2; Length 199311;  
 Best Local Similarity 100.0%; Pred. No. 7.4e-06;  
 Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 440 AAACCACTCTTTTTTTTTTTTTTTTTTTTTTTTTTTT 470  
 Dd 47271 AAACCACTCTTTTTTTTTTTTTTTTTTTTTTTTTTTT 47241

# RESULT 12

AC093465  
 LOCUS AC093465 240682 bp DNA linear HTG 26-AUG-2001  
 DEFINITION Mus musculus clone RP23-256C1, WORKING DRAFT SEQUENCE, 11 unordered  
 pieces.  
 AC093465  
 AC093465.1 GI:15290911  
 HTG: HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.  
 KEYWORDS house mouse.  
 SOURCE Mus musculus  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
 1 (bases 1 to 240682)  
 Unpublished  
 2 (bases 1 to 240682)  
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,  
 Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhgalter,B.,  
 Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,  
 Choquel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,  
 Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,  
 Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,  
 Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-pierre,N.,  
 Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,  
 Jones,C., Kanat,A., Karatas,A., Kells,C., LaReccque,K.,  
 Lamarez,R., Landers,T., Lechoczky,J., Levine,R., Liu,G.,  
 MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,  
 McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,  
 Menes,L., Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C.,  
 Niba,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,  
 Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,  
 Raymond,C., Retta,R., Riback,M., Riley,R., Rise,C., Rogov,P.,  
 Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupbach,R.,  
 Seaman,S., Severy,E., Spencer,B., Stange-Thomann,N., Stojanovic,N.,  
 Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,  
 Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,

Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,  
 Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.  
 Direct Submission  
 Submitted (26-AUG-2001) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 All repeats were identified using RepeatMasker:  
 Sait,A.F.A. & Green,P. (1996-1997).  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html  
 -----  
 Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence\_submissions@genome.wi.mit.edu  
 -----  
 Project Information  
 Center project name: L13279  
 Center clone name: 256\_C1  
 -----  
 Summary Statistics  
 Sequencing vector: Plasmid; n/a; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.960731  
 Consensus quality: 236849 bases at least Q40  
 Consensus quality: 238347 bases at least Q30  
 Consensus quality: 238907 bases at least Q20  
 Insert size: 242000; agarose-fp  
 Quality coverage: 11.0 in Q20 bases; agarose-fp  
 Quality coverage: 11.1 in Q20 bases; sum-of-contigs

-----  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 11 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \* 1 12397: contig of 12397 bp in length  
 \* 12398 12497: gap of 100 bp  
 \* 12498 13891: contig of 1394 bp in length  
 \* 13892 13991: gap of 100 bp  
 \* 13992 15867: contig of 1876 bp in length  
 \* 15868 15967: gap of 100 bp  
 \* 15968 18460: contig of 2493 bp in length  
 \* 18461 18560: gap of 100 bp  
 \* 18561 28925: contig of 10365 bp in length  
 \* 28926 29025: gap of 100 bp  
 \* 29026 48714: contig of 19688 bp in length  
 \* 48715 48814: gap of 100 bp  
 \* 48815 112226: contig of 63412 bp in length  
 \* 112227 112326: gap of 100 bp  
 \* 112327 151662: contig of 39336 bp in length  
 \* 151663 151762: gap of 100 bp  
 \* 151763 180086: contig of 28324 bp in length  
 \* 180087 180186: gap of 100 bp  
 \* 180187 219865: contig of 39679 bp in length  
 \* 219866 219965: gap of 100 bp  
 \* 219966 240682: contig of 20717 bp in length.  
 Location/Qualifiers  
 1..240682  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /clone="RP23-256C1"  
 /clone\_lib="RPC1-23 Female Mouse BAC"  
 1..12397  
 /note="assembly\_fragment  
 clone\_end:SP6  
 vector\_side:left"  
 12498..13891  
 /note="assembly\_fragment"  
 13992..15867  
 /note="assembly\_fragment"  
 15968..18460  
 /note="assembly\_fragment"

-----  
 misc\_feature  
 misc\_feature  
 misc\_feature  
 misc\_feature

```

misc_feature 18561..28925
/note="assembly_fragment"
misc_feature 29026..48714
/note="assembly_fragment"
misc_feature 48815..112226
/note="assembly_fragment"
misc_feature 112327..151662
/note="assembly_fragment"
misc_feature 151763..180086
/note="assembly_fragment"
misc_feature 180187..219865
/note="assembly_fragment"
misc_feature 219966..240682
/note="assembly_fragment"
clone_end:T7
vector_side:right"
BASE COUNT 66652 a 53178 c 53976 g 65866 t 1010 others
ORIGIN

Query Match      4.3%; Score 31; DB 2; Length 240682;
Best Local Similarity 100.0%; Pred. No. 7.3e-06;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 445 ACCTCTTTTTTTTTTTTTTTTTTTTTTTTGGCTT 475
|TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT|
Db 97793 ACCTCTTTTTTTTTTTTTTTTTTTTTTTTGGCTT 97823

RESULT 13
LOCUS AC101852 65400 bp DNA linear HTG 23-NOV-2001
DEFINITION Mus musculus clone RP23-279P3, LOW-PASS SEQUENCE SAMPLING.
ACCESSION AC101852
VERSION AC101852.1 GI:17060627
KEYWORDS HTG; HTGS_PHASE0.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 65400)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Mus musculus, clone RP23-279P3
2 (bases 1 to 65400)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavskiy,L., Boukhgalter,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
Choepe,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Kamat,A., Karatas,A., Kells,C., LaRoque,K.,
Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Melidrim,J.,
Meneus,L., Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuback,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
Center: Whitehead Institute/MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: LI7649
Center clone name: 279_P3
-----
* NOTE: This record contains 82 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*
1 681: contig of 681 bp in length
682 781: gap of 100 bp
782 1474: contig of 693 bp in length
1475 1574: gap of 100 bp
1575 2275: contig of 701 bp in length
2276 2375: gap of 100 bp
2376 3097: contig of 722 bp in length
3098 3197: gap of 100 bp
3198 3898: contig of 701 bp in length
3899 3998: gap of 100 bp
3999 4713: contig of 715 bp in length
4714 4813: gap of 100 bp
4814 5516: contig of 703 bp in length
5517 5616: gap of 100 bp
5617 6318: contig of 702 bp in length
6319 6418: gap of 100 bp
6419 7112: contig of 694 bp in length
7113 7212: gap of 100 bp
7213 7919: contig of 707 bp in length
7920 8019: gap of 100 bp
8020 8719: contig of 700 bp in length
8720 8819: gap of 100 bp
8820 9511: contig of 692 bp in length
9512 9611: gap of 100 bp
9612 10326: contig of 715 bp in length
10327 10426: gap of 100 bp
10427 11119: contig of 693 bp in length
11120 11219: gap of 100 bp
11220 11936: contig of 717 bp in length
11937 12036: gap of 100 bp
12037 12730: contig of 694 bp in length
12731 12830: gap of 100 bp
12831 13523: contig of 693 bp in length
13524 13623: gap of 100 bp
13624 14325: contig of 702 bp in length
14326 14425: gap of 100 bp
14426 15121: contig of 696 bp in length
15122 15221: gap of 100 bp
15222 15921: contig of 700 bp in length
15922 16021: gap of 100 bp
16022 16714: contig of 693 bp in length
16715 16814: gap of 100 bp
16815 17511: contig of 697 bp in length
17512 17611: gap of 100 bp
17612 18298: contig of 687 bp in length
18299 18398: gap of 100 bp
18399 19103: contig of 705 bp in length
19104 19203: gap of 100 bp
19204 19898: contig of 695 bp in length
19899 19998: gap of 100 bp
19999 20708: contig of 710 bp in length
20709 20808: gap of 100 bp
20809 21570: contig of 762 bp in length

```

\* 21571 21670: gap of 100 bp  
\* 21671 22354: contig of 684 bp in length  
\* 22355 22454: gap of 100 bp  
\* 22455 23154: contig of 700 bp in length  
\* 23155 23254: gap of 100 bp  
\* 23255 23945: contig of 691 bp in length  
\* 23946 24045: gap of 100 bp  
\* 24046 24725: contig of 680 bp in length  
\* 24726 24825: gap of 100 bp  
\* 24826 25535: contig of 710 bp in length  
\* 25536 25635: gap of 100 bp  
\* 25636 26323: contig of 688 bp in length  
\* 26324 26423: gap of 100 bp  
\* 26424 27143: contig of 720 bp in length  
\* 27144 27243: gap of 100 bp  
\* 27244 27956: contig of 713 bp in length  
\* 27957 28056: gap of 100 bp  
\* 28057 28758: contig of 702 bp in length  
\* 28759 28858: gap of 100 bp  
\* 28859 29546: contig of 688 bp in length  
\* 29547 29646: gap of 100 bp  
\* 29647 30316: contig of 670 bp in length  
\* 30317 30416: gap of 100 bp  
\* 30417 31117: contig of 701 bp in length  
\* 31118 31217: gap of 100 bp  
\* 31218 31918: contig of 701 bp in length  
\* 31919 32018: gap of 100 bp  
\* 32019 32700: contig of 682 bp in length  
\* 32701 32800: gap of 100 bp  
\* 32801 33491: contig of 691 bp in length  
\* 33492 33591: gap of 100 bp  
\* 33592 34292: contig of 701 bp in length  
\* 34293 34392: gap of 100 bp  
\* 34393 35090: contig of 698 bp in length  
\* 35091 35190: gap of 100 bp  
\* 35191 35895: contig of 705 bp in length  
\* 35896 35995: gap of 100 bp  
\* 35996 36711: contig of 716 bp in length  
\* 36712 36811: gap of 100 bp  
\* 36812 37518: contig of 707 bp in length  
\* 37519 37618: gap of 100 bp  
\* 37619 38317: contig of 699 bp in length  
\* 38318 38417: gap of 100 bp  
\* 38418 39125: contig of 708 bp in length  
\* 39126 39225: gap of 100 bp  
\* 39226 39919: contig of 694 bp in length  
\* 40020 40019: gap of 100 bp  
\* 40020 40718: contig of 699 bp in length  
\* 40719 40818: gap of 100 bp  
\* 40819 41513: contig of 695 bp in length  
\* 41514 41613: gap of 100 bp  
\* 41614 42307: contig of 694 bp in length  
\* 42308 42407: gap of 100 bp  
\* 42408 43099: contig of 692 bp in length  
\* 43100 43199: gap of 100 bp  
\* 43200 43905: contig of 706 bp in length  
\* 43906 44005: gap of 100 bp  
\* 44006 44716: contig of 711 bp in length  
\* 44717 44816: gap of 100 bp  
\* 44817 45478: contig of 662 bp in length  
\* 45479 45578: gap of 100 bp  
\* 45579 46288: contig of 710 bp in length  
\* 46289 46388: gap of 100 bp  
\* 46389 47080: contig of 692 bp in length  
\* 47081 47180: gap of 100 bp  
\* 47181 47877: contig of 697 bp in length  
\* 47878 47977: gap of 100 bp  
\* 47978 48674: contig of 697 bp in length  
\* 48675 48774: gap of 100 bp  
\* 48775 49470: contig of 696 bp in length  
\* 49471 49570: gap of 100 bp  
\* 49571 50263: contig of 693 bp in length  
\* 50264 50363: gap of 100 bp

\* 50364 51056: contig of 693 bp in length  
\* 51057 51156: gap of 100 bp  
\* 51157 51863: contig of 707 bp in length  
\* 51864 51963: gap of 100 bp  
\* 51964 52673: contig of 710 bp in length  
\* 52674 52773: gap of 100 bp  
\* 52774 53488: contig of 715 bp in length  
\* 53489 53588: gap of 100 bp  
\* 53589 54302: contig of 714 bp in length  
\* 54303 54402: gap of 100 bp  
\* 54403 55110: contig of 708 bp in length  
\* 55111 55210: gap of 100 bp

Query Match 4.1%; Score 30; DB 2; Length 65400;  
Best Local Similarity 100.0%; Pred. No. 2.7e-05;  
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 449 TTTTGTGCTTCTTTTGGCTTCTTT 478

Db 47345 TTTTGTGCTTCTTTTGGCTTCTTT 47374

#### RESULT 14

##### LOCUS

AC102249 Mus musculus clone RP24-160D21, LOW-PASS SEQUENCE SAMPLING. 67166 bp DNA linear HTG 23-NOV-2001

##### DEFINITION

AC102249

##### ACCESSION

AC102249.1 GI:17061335

##### VERSION

HTG: HTGS\_PHASE0.

##### KEYWORDS

house mouse.

##### SOURCE

Mus musculus

##### ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus;

Birren, B., Linton, L., Nusbaum, C. and Lander, E.

1 (bases 1 to 67166)

Mus musculus, clone RP24-160D21

Unpublished

2 (bases 1 to 67166)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,

Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhgalter, B.,

Brown, A., Camarata, J., Campolano, A., Chang, J., Chazaro, B.,

Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,

Cooke, P., DeArellano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S.,

Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S.,

Ginde, S., Gord, S., Govett, M., Graham, L., Grand-Pierre, N.,

Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliiev, I., Johnson, R.,

Jones, C., Kamat, A., Karatas, A., Kells, C., LaRoque, K.,

Lamarez, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,

Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,

McCarthy, M., McSwan, P., McKernan, K., McSheeters, R., Meldrum, J.,

Menius, L., Milhova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C.,

Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D.,

Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollard, V.,

Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,

Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuback, R.,

Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,

Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,

Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,

Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,

Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission

Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WBIR

Web site: http://www-seq.wi.mit.edu

Contact: sequence-submissions@genome.wi.mit.edu

----- Project Information

Center project name: L18261

Center clone name: 160\_D\_21

\*\*\*\*\*  
\* NOTE: This record contains 87 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

1 638: contig of 638 bp in length  
\* 639 738: gap of 100 bp  
\* 739 1417: contig of 679 bp in length  
\* 1418 1517: gap of 100 bp  
\* 1518 2201: contig of 684 bp in length  
\* 2202 2301: gap of 100 bp  
\* 2302 2985: contig of 684 bp in length  
\* 2986 3085: gap of 100 bp  
\* 3086 3747: contig of 662 bp in length  
\* 3748 3847: gap of 100 bp  
\* 3848 4525: contig of 678 bp in length  
\* 4526 4625: gap of 100 bp  
\* 4626 5312: contig of 687 bp in length  
\* 5313 5412: gap of 100 bp  
\* 5413 6108: contig of 696 bp in length  
\* 6109 6208: gap of 100 bp  
\* 6209 6896: contig of 688 bp in length  
\* 6897 6996: gap of 100 bp  
\* 6997 7685: contig of 689 bp in length  
\* 7686 7785: gap of 100 bp  
\* 7786 8466: contig of 681 bp in length  
\* 8467 8566: gap of 100 bp  
\* 8567 9248: contig of 682 bp in length  
\* 9249 9348: gap of 100 bp  
\* 9349 10021: contig of 673 bp in length  
\* 10022 10121: gap of 100 bp  
\* 10122 10685: contig of 564 bp in length  
\* 10686 10785: gap of 100 bp  
\* 10786 11453: contig of 668 bp in length  
\* 11454 11553: gap of 100 bp  
\* 11554 12236: contig of 683 bp in length  
\* 12237 12336: gap of 100 bp  
\* 12337 13007: contig of 671 bp in length  
\* 13008 13107: gap of 100 bp  
\* 13108 13782: contig of 675 bp in length  
\* 13783 13882: gap of 100 bp  
\* 13883 14556: contig of 674 bp in length  
\* 14557 14656: gap of 100 bp  
\* 14657 15345: contig of 689 bp in length  
\* 15346 15445: gap of 100 bp  
\* 15446 16116: contig of 671 bp in length  
\* 16117 16216: gap of 100 bp  
\* 16217 16888: contig of 672 bp in length  
\* 16889 16988: gap of 100 bp  
\* 16989 17667: contig of 679 bp in length  
\* 17668 17767: gap of 100 bp  
\* 17768 18450: contig of 683 bp in length  
\* 18451 18550: gap of 100 bp  
\* 18551 19204: contig of 654 bp in length  
\* 19205 19304: gap of 100 bp  
\* 19305 19971: contig of 667 bp in length  
\* 19972 20071: gap of 100 bp  
\* 20072 20753: contig of 682 bp in length  
\* 20754 20853: gap of 100 bp  
\* 20854 21526: contig of 673 bp in length  
\* 21527 21626: gap of 100 bp  
\* 21627 22305: contig of 679 bp in length  
\* 22306 22405: gap of 100 bp  
\* 22406 23071: contig of 666 bp in length  
\* 23072 23171: gap of 100 bp

\* 23172 23850: contig of 679 bp in length  
\* 23851 23950: gap of 100 bp  
\* 23951 24616: contig of 666 bp in length  
\* 24617 24716: gap of 100 bp  
\* 24717 25386: contig of 670 bp in length  
\* 25387 25486: gap of 100 bp  
\* 25487 26176: contig of 690 bp in length  
\* 26177 26276: gap of 100 bp  
\* 26277 26948: contig of 672 bp in length  
\* 26949 27048: gap of 100 bp  
\* 27049 27699: contig of 651 bp in length  
\* 27700 27799: gap of 100 bp  
\* 27800 28468: contig of 669 bp in length  
\* 28469 28568: gap of 100 bp  
\* 28569 29233: contig of 665 bp in length  
\* 29234 29333: gap of 100 bp  
\* 29334 30005: contig of 672 bp in length  
\* 30006 30105: gap of 100 bp  
\* 30106 30767: contig of 662 bp in length  
\* 30768 30867: gap of 100 bp  
\* 30868 31523: contig of 656 bp in length  
\* 31524 31623: gap of 100 bp  
\* 31624 32296: contig of 673 bp in length  
\* 32297 32396: gap of 100 bp  
\* 32397 33086: contig of 690 bp in length  
\* 33087 33186: gap of 100 bp  
\* 33187 33856: contig of 670 bp in length  
\* 33857 33956: gap of 100 bp  
\* 33957 34610: contig of 654 bp in length  
\* 34611 34710: gap of 100 bp  
\* 34711 35382: contig of 672 bp in length  
\* 35383 35482: gap of 100 bp  
\* 35483 36143: contig of 661 bp in length  
\* 36144 36243: gap of 100 bp  
\* 36244 36911: contig of 668 bp in length  
\* 36912 37011: gap of 100 bp  
\* 37012 37678: contig of 667 bp in length  
\* 37679 37778: gap of 100 bp  
\* 37779 38447: contig of 669 bp in length  
\* 38448 38547: gap of 100 bp  
\* 38548 39220: contig of 673 bp in length  
\* 39221 39320: gap of 100 bp  
\* 39321 40006: contig of 686 bp in length  
\* 40007 40106: gap of 100 bp  
\* 40107 40795: contig of 689 bp in length  
\* 40796 40895: gap of 100 bp  
\* 40896 41566: contig of 671 bp in length  
\* 41567 41666: gap of 100 bp  
\* 41667 42363: contig of 697 bp in length  
\* 42364 42463: gap of 100 bp  
\* 42464 43127: contig of 664 bp in length  
\* 43128 43227: gap of 100 bp  
\* 43228 43918: contig of 691 bp in length  
\* 43919 44018: gap of 100 bp  
\* 44019 44703: contig of 685 bp in length  
\* 44704 44803: gap of 100 bp  
\* 44804 45468: contig of 665 bp in length  
\* 45469 45568: gap of 100 bp  
\* 45569 46238: contig of 670 bp in length  
\* 46239 46338: gap of 100 bp  
\* 46339 47012: contig of 674 bp in length  
\* 47013 47112: gap of 100 bp  
\* 47113 47783: contig of 671 bp in length  
\* 47784 47883: gap of 100 bp  
\* 47884 48563: contig of 680 bp in length  
\* 48564 48663: gap of 100 bp  
\* 48664 49331: contig of 668 bp in length  
\* 49332 49431: gap of 100 bp  
\* 49432 50108: contig of 677 bp in length  
\* 50109 50208: gap of 100 bp  
\* 50209 50869: contig of 661 bp in length  
\* 50870 50969: gap of 100 bp  
\* 50970 51654: contig of 685 bp in length





26177 26276: gap of 100 bp  
26277 26948: contig of 672 bp in length  
26949 27048: gap of 100 bp  
27049 27699: contig of 651 bp in length  
27700 27799: gap of 100 bp  
27800 28468: contig of 669 bp in length  
28469 28568: gap of 100 bp  
28569 29233: contig of 665 bp in length  
29234 29333: gap of 100 bp  
29334 30005: contig of 672 bp in length  
30006 30105: gap of 100 bp  
30106 30767: contig of 662 bp in length  
30768 30867: gap of 100 bp  
30868 31523: contig of 656 bp in length  
31524 31623: gap of 100 bp  
31624 32296: contig of 673 bp in length  
32297 32396: gap of 100 bp  
32397 33086: contig of 690 bp in length  
33087 33186: gap of 100 bp  
33187 33856: contig of 670 bp in length  
33857 33956: gap of 100 bp  
33957 34610: contig of 654 bp in length  
34611 34710: gap of 100 bp  
34711 35382: contig of 672 bp in length  
35383 35482: gap of 100 bp  
35483 36143: contig of 661 bp in length  
36144 36243: gap of 100 bp  
36244 36911: contig of 668 bp in length  
36912 37011: gap of 100 bp  
37012 37678: contig of 667 bp in length  
37679 37778: gap of 100 bp  
37779 38447: contig of 669 bp in length  
38448 38547: gap of 100 bp  
38548 39220: contig of 673 bp in length  
39221 39320: gap of 100 bp  
39321 40006: contig of 686 bp in length  
40007 40106: gap of 100 bp  
40107 40795: contig of 689 bp in length  
40796 40895: gap of 100 bp  
40896 41566: contig of 671 bp in length  
41567 41666: gap of 100 bp  
41667 42363: contig of 697 bp in length  
42364 42463: gap of 100 bp  
42464 43127: contig of 664 bp in length  
43128 43227: gap of 100 bp  
43228 43918: contig of 691 bp in length  
43919 44018: gap of 100 bp  
44019 44703: contig of 685 bp in length  
44704 44803: gap of 100 bp  
44804 45468: contig of 665 bp in length  
45469 45568: gap of 100 bp  
45569 46238: contig of 670 bp in length  
46239 46338: gap of 100 bp  
46339 47012: contig of 674 bp in length  
47013 47112: gap of 100 bp  
47113 47783: contig of 671 bp in length  
47784 47883: gap of 100 bp  
47884 48563: contig of 680 bp in length  
48564 48663: gap of 100 bp  
48664 49331: contig of 668 bp in length  
49332 49431: gap of 100 bp  
49432 50108: contig of 677 bp in length  
50109 50208: gap of 100 bp  
50209 50869: contig of 661 bp in length  
50870 50969: gap of 100 bp  
50970 51654: contig of 685 bp in length  
51655 51754: gap of 100 bp  
51755 52429: contig of 675 bp in length  
52430 52529: gap of 100 bp  
52530 53218: contig of 689 bp in length  
53219 53318: gap of 100 bp

Query Match 4.1%; Score 30; DB 2; Length 67166;

Best Local Similarity 100.0%; Pred. No. 2.7e-05;  
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 448 CTTTTTTTTTTTTTTTTTTGGCTTCT 477  
Db 35316 CTTTTTTTTTTTTTTTTTTGGCTTCT 35287

Search completed: August 3, 2002, 01:41:11  
Job time: 40641 sec

us-08-973-363-1.oli.rge

Mon Aug 5 11:51:32 2002

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 2, 2002, 16:53:35 ; Search time 949.74 Seconds  
(without alignments)  
1307.020 Million cell updates/sec

Title: US-08-973-363-1

Perfect score: 723

Sequence: 1 CCCGTCGAGGTTTCAAGG.....CTTGAAACCTCCGACCGG 723

Scoring table: OLIGO\_NUC

Gapop 60.0 , Gapext 60.0

Searched: 1736436 seqs, 858457221 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

- N\_Geneseq\_032802:\*
- 1: /SIDSI/gcgdata/hold-geneq/geneq-emb1/NA1980.DAT:\*
  - 2: /SIDSI/gcgdata/hold-geneq/geneq-emb1/NA1981.DAT:\*
  - 3: /SIDSI/gcgdata/hold-geneq/geneq-emb1/NA1982.DAT:\*
  - 4: /SIDSI/gcgdata/hold-geneq/geneq-emb1/NA1983.DAT:\*
  - 5: /SIDSI/gcgdata/hold-geneq/geneq-emb1/NA1984.DAT:\*
  - 6: /SIDSI/gcgdata/hold-geneq/geneq-emb1/NA1985.DAT:\*
  - 7: /SIDSI/gcgdata/hold-geneq/geneq-emb1/NA1986.DAT:\*
  - 8: /SIDSI/gcgdata/hold-geneq/geneq-emb1/NA1987.DAT:\*
  - 9: /SIDSI/gcgdata/hold-geneq/geneq-emb1/NA1988.DAT:\*
  - 10: /SIDSI/gcgdata/hold-geneq/geneq-emb1/NA1989.DAT:\*
  - 11: /SIDSI/gcgdata/hold-geneq/geneq-emb1/NA1990.DAT:\*
  - 12: /SIDSI/gcgdata/hold-geneq/geneq-emb1/NA1991.DAT:\*
  - 13: /SIDSI/gcgdata/hold-geneq/geneq-emb1/NA1992.DAT:\*
  - 14: /SIDSI/gcgdata/hold-geneq/geneq-emb1/NA1993.DAT:\*
  - 15: /SIDSI/gcgdata/hold-geneq/geneq-emb1/NA1994.DAT:\*
  - 16: /SIDSI/gcgdata/hold-geneq/geneq-emb1/NA1995.DAT:\*
  - 17: /SIDSI/gcgdata/hold-geneq/geneq-emb1/NA1996.DAT:\*
  - 18: /SIDSI/gcgdata/hold-geneq/geneq-emb1/NA1997.DAT:\*
  - 19: /SIDSI/gcgdata/hold-geneq/geneq-emb1/NA1998.DAT:\*
  - 20: /SIDSI/gcgdata/hold-geneq/geneq-emb1/NA1999.DAT:\*
  - 21: /SIDSI/gcgdata/hold-geneq/geneq-emb1/NA2000.DAT:\*
  - 22: /SIDSI/gcgdata/hold-geneq/geneq-emb1/NA2001A.DAT:\*
  - 23: /SIDSI/gcgdata/hold-geneq/geneq-emb1/NA2001B.DAT:\*
  - 24: /SIDSI/gcgdata/hold-geneq/geneq-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Query	Score	Match	Length	ID	Description
1	723	100.0	723	18	AAT42755	Great tit CHD-W ge
2	29	4.0	429	22	AAI88877	Human polynucleoti
3	29	4.0	755	11	AAQ03970	Clone P25b. Mus m
4	29	4.0	1564	22	AAI64562	Human polypeptide-
5	29	4.0	11630	22	AAS02362	Galactosyl transfe
6	29	4.0	27048	22	AAS02354	Porcine Gal alpha(
7	28	3.9	450	21	AAH45071	Human secreted exp
8	28	3.9	499	21	AAZ97350	Human prostatic can
9	28	3.9	626	22	ABAI7747	Human nervous syst

10	28	3.9	626	22	ABA21398	Human nervous syst
c 11	28	3.9	761	22	AAL11085	Human breast cancer
c 12	28	3.9	1008	24	ABA02499	Maize pathogen res
13	28	3.9	1086	22	AAK70887	Human immune/haema
14	28	3.9	1086	22	AAK70888	Human immune/haema
15	28	3.9	1086	22	AAK94102	Human immune/haema
16	28	3.9	1086	22	AAK94103	Human immune/haema
c 17	28	3.9	3356	24	AAK22731	Human cDNA encodin
c 18	28	3.9	5860	24	AAS61372	Human gene regulat
c 19	28	3.9	9316	22	AAS28637	Genomic sequence #
20	28	3.9	12595	22	AAS42100	Genomic sequence #
c 21	28	3.9	30826	22	ABA07345	Human pancreatic c
c 22	28	3.9	30826	22	AAS32772	Human genomic DNA
23	28	3.9	465237	24	ABA90193	Human oestrogen re
24	27	3.7	160	22	AAL24600	Human breast cancer
c 25	27	3.7	225	22	AAK85883	Human immune/haema
26	27	3.7	269	22	AAL10304	Human breast cancer
27	27	3.7	342	22	AAK13657	Human breast cancer
28	27	3.7	346	22	AAL22526	Human breast cancer
29	27	3.7	362	22	AAL13702	Human breast cancer
30	27	3.7	362	22	AAL22571	Human breast cancer
31	27	3.7	399	22	AAI19737	Human breast cancer
32	27	3.7	404	22	AAI79997	Human polynucleoti
33	27	3.7	415	22	AAL19638	Human breast cancer
c 34	27	3.7	486	22	AAK74174	Human immune/haema
c 35	27	3.7	486	22	AAK74175	Human immune/haema
c 36	27	3.7	549	19	AAV38759	CDNA ofSPI7-549 en
37	27	3.7	549	19	AAV38760	Complementary stra
c 38	27	3.7	578	21	AAK76861	Human ORF2416
c 39	27	3.7	622	21	AAI16179	Human colon cancer
40	27	3.7	717	22	AAL19884	Human breast cancer
41	27	3.7	767	22	AAL24165	Human breast cancer
c 42	27	3.7	802	22	AAK87192	Human immune/haema
c 43	27	3.7	1089	20	AAV84474	Human secreted pro
c 44	27	3.7	1089	22	ABA83257	Human secreted pro
c 45	27	3.7	1085	22	AAS29609	Human endocrine po

## ALIGNMENTS

RESULT 1  
AAT42755  
ID AAT42755 standard; DNA: 723 BP.

XX AAT42755;

XX 12-MAR-1997 (first entry)

XX Great tit CHD-W gene probe (pGT-W insert).

XX Bird; sex determination; chromodomain-Helicase-DNA binding 1 Avian;

XX CHD-W; chromodomain-Helicase-DNA binding on the W chromosome; ss.

XX Parus major.

XX WO9699505-A1.

XX 12-DEC-1996.

XX 05-JUN-1996; 96WO-G801341.

XX 06-JUN-1995; 95GB-0011439.

XX (ISIS-) ISIS INNOVATION LTD.

XX Griffiths R, Tiwari B;

XX WPI; 1997-043127/04.

XX Avian chromodomain-Helicase-DNA binding genes determine sex in birds - used for sex determ. and to control sex of progeny



```
XX DE Clone P25b.
XX DE
XX KW Ehrlich ascites tumour; protein P25; ss.
XX OS Mus musculus.
XX FH Location/Qualifiers
XX FT 91..693
XX FT CDS
XX FT /tag= a
XX FT /label=P25 protein
XX
XX PN DD273071-A.
XX
XX PD 01-NOV-1989.
XX
XX PF 17-JUN-1988; 88DD-0168546.
XX
XX PR 17-JUN-1988; 88DD-0316854.
XX
XX PA (DEAK ) AKAD WISSENSCHAFT DDR.
XX
XX PI Gaestel M, Gross B, Benndorf R, Bielka H;
XX
XX DR WPI; 1990-108215/15.
XX DR P-PSDB; AAR05780.
XX
XX PT New protein P25 prodn. - using P25 cDNA synthesised in vitro from ehrlich
XX PT ascites tumour cell total mRNA
XX
XX PS Disclosure; fig 1; 6pp; German.
XX
XX CC Complementary (c) DNA was transcribed from Ehrlich ascites tumour poly-A
XX CC mRNA and subsequent hybridisation of this with oligonucleotide probes
XX CC resulted in 2 clones of which 1 (P25b) comprised this sequence. The
XX CC other clone (P25a) was altered at the translation start site by
XX CC site-directed mutagenesis (at Nde I site). This was then used to
XX CC transform host cells and the prod. is useful in eg medical research. See
XX CC also AAQ03969.
XX
XX SQ Sequence 755 BP; 181 A; 228 C; 212 G; 133 T; 1 other;

Query Match 4.0%; Score 29; DB 11; Length 755;
Best Local Similarity 100.0%; Pred. No. 0.019;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 449 TTTTTCCTTTTTCCTTTTTCCTTTCT 477
DB 742 TTTTTCCTTTTTCCTTTTTCCTTTCT 714

RESULT 4
AAI64562
ID AAI64562 standard; cDNA; 1564 BP.
AC AAI64562;
XX
XX DT 04-DEC-2001 (first entry)
XX
XX DE Human polypeptide-tyrosinase 14 encoding cDNA.
XX
XX KW Human; polypeptide-tyrosinase 14; cancer; human immunodeficiency virus;
XX KW HIV; infection; ss.
XX
XX OS Homo sapiens.
XX
XX FH Location/Qualifiers
XX FT 1042..1428
XX FT CDS
XX FT /tag= a
XX FT /product= "polypeptide-tyrosinase 14"
XX FT /note= "claimed in claim 6"
XX FT
XX FT
XX FT
```

```
PN CNI300743-A.
XX
XX PD 27-JUN-2001.
XX
XX PF 21-DEC-1999; 99CN-0125656.
XX
XX PR 21-DEC-1999; 99CN-0125656.
XX
XX PA (UYFU-) UNIV FUDAN.
XX
XX PI Mao Y, Xie Y;
XX
XX DR WPI; 2001-530454/59.
XX DR P-PSDB; AAG78154.
XX
XX PT Polypeptide-tyrosinase 14 for treating e.g. cancer and human
XX PT immunodeficiency virus (HIV) infection.
XX
XX PS Claim 6; Page 23-24 (Disclosure); 31pp; Chinese.
XX
XX CC The invention relates to the human polypeptide-tyrosinase 14, the
XX CC polynucleotide for coding it and the application of the polypeptide in
XX CC treating diseases including cancer, human immunodeficiency virus (HIV)
XX CC infection.
XX
XX SQ Sequence 1564 BP; 376 A; 377 C; 380 G; 431 T; 0 other;

Query Match 4.0%; Score 29; DB 22; Length 1564;
Best Local Similarity 100.0%; Pred. No. 0.017;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 443 CCACCTCTTTTTCCTTTTTCCTTTTCG 471
DB 884 ccactctttttttttttttttttttg 912

RESULT 5
AA502362
ID AA502362 standard; DNA; 11630 BP.
AC AA502362;
XX
XX DT 12-SEP-2001 (first entry)
XX
XX DE Galactosyl transferase gene targeting vector.
XX
XX KW Gai alpha(1,3) galactosyl transferase; gene targeting;
XX KW transgenic animal; transplant rejection; immunomodulation;
XX KW systemic lupus erythematosus; immune-haemolytic anaemia;
XX KW collision intron targeting construct; rheumatoid arthritis; ds.
XX
XX OS Synthetic.
XX OS Sus scrofa.
XX
XX FH Location/Qualifiers
XX FT 235..260
XX FT primer_bind
XX FT /tag= a
XX FT /note= "PCR primer as displayed in AA502349"
XX FT complement (4827..4851)
XX FT primer_bind
XX FT /tag= b
XX FT /note= "PCR primer as displayed in AA502350"
XX FT 4851..4852
XX FT misc_feature
XX FT /tag= c
XX FT /note= "Site of insertion of the puromycin/bovine
XX FT growth hormone poly-A signal sequence"
XX FT 4852..4875
XX FT primer_bind
XX FT /tag= d
XX FT /note= "PCR primer as displayed in AA502351"
XX FT complement (5911..6935)
XX FT primer_bind
XX FT /tag= e
XX FT /note= "PCR primer as displayed in AA502352"
XX FT
XX FT
```





PI Fodor WL, Ramscoondar JJ;  
XX WPI; 2001-266147/27.  
XX  
XX Modulating the expression of a eukaryotic gene in a cell, involves  
PT transfecting the cell with a nucleic acid construct that disrupts at  
PT least a portion of the DNA sequence of the gene to be modulated -  
XX  
XX Example 1; Fig 1; 86pp; English.  
XX  
XX The sequence represents Porcine Gal alpha(1,3) galactosyl transferase  
CC introns 3-8, a gene used to make a gene targeting construct of the  
CC invention. The invention relates to a method for modulating the  
CC expression of a eukaryotic gene in a cell. The method involves  
CC transfecting the cell with a nucleic acid construct. The construct  
CC contains 2 nucleotide sequences which are portions of one or more  
CC introns of the eukaryotic gene, and a sequence encoding a selectable  
CC marker. The marker sequence is integrated into the gene sequence, so  
CC that expression of the marker results in modulation of the expression  
CC of the gene. The construct is useful for making a transgenic mammal.  
CC The construct is useful for reducing transplant rejection by harvesting  
CC cells, tissue, or organs from the offspring of the transgenic mammal,  
CC and transplanting the harvested cells, tissue, or organs into a patient  
CC in need. The knockout mammals produced by the method are useful for  
CC screening drugs for immunomodulation (e.g. for systemic lupus  
CC erythematosus, rheumatoid arthritis and immune-haemolytic anaemia) and  
CC for producing proteins of interest.  
XX  
XX Sequence 27048 BP; 6964 A; 5958 C; 6323 G; 7365 T; 438 other;

Query Match 4.0%; Score 29; DB 22; Length 27048;  
Best Local Similarity 100.0%; Pred. No. 0.011;  
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 447 TCTTTTTTTTTTTTTTTTTTTTTTTGGCTT 475  
|||||  
Db 10987 tcttttttttttttttttttttttggctt 11015

RESULT 7  
AAA45071  
ID AAA45071 standard; cDNA; 450 BP.  
XX  
XX AAA45071:  
XX  
XX 21-AUG-2000 (first entry)  
XX Human secreted expressed sequence tag SEQ ID NO:1646.  
XX  
XX Human; mouse; chicken; rat; secreted expressed sequence tag; SEST;  
XX expressed sequence tag; EST; probe; chemotactic; proliferative;  
XX immunomodulatory; haemopoietic; chemokine; analgesic; haemostatic;  
XX thrombolytic; antiinflammatory; cytostatic; antibacterial; antifungal;  
XX antiviral; antidiabetic; antiaesthetic; vulnary; antiparkinsonian;  
XX antitumor; osteoprotective; neuroprotective; nootropic; antipsoriatic;  
XX cerebroprotective; anticonvulsant; antidepressant; gene therapy;  
XX vaccine; autoimmune disorder; multiple sclerosis; allergic condition;  
XX insulin dependent diabetes; asthma; myeloid cell deficiency; ulcer;  
XX lymphoid cell deficiency; burn; osteoporosis; osteoarthritis;  
XX central nervous system disorder; Alzheimer's disease; stroke;  
XX Parkinson's disease; Huntington's disease; coagulation disorder;  
XX haemophilia; thrombosis; inflammatory disorder; Crohn's disease;  
XX tumour; infection; depression; psoriasis; ss.  
XX  
XX Homo sapiens.  
XX  
XX WO200021991-A1.  
XX  
XX 20-APR-2000.  
XX  
XX 15-OCT-1999; 99WO-US24206.

PR 15-OCT-1998; 98US-0104436.  
XX (GENY) GENETICS INST INC.  
XX  
XX Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;  
PI Merberg D, Treacy M, Bowman MR;  
XX WPI; 2000-317938/27.  
XX  
XX Isolated polynucleotides, and encoded proteins, comprising secreted  
PT expressed sequence tags (SESTs), useful for treating various disorders -  
PT such as autoimmune, infectious, and central nervous system disorders -  
XX  
XX Claim 1; Page 630; 803pp; English.  
XX  
XX AAA43426 to AAA45925 represent specifically claimed secreted expressed  
CC sequence tags (SESTs), isolated from human, mouse, chicken and rat  
CC tissue sources. The SESTs can have a range of activities depending on  
CC the tissues they were isolated from. The activities include:  
CC chemotactic; proliferative; immunomodulatory; haemopoietic;  
CC chemokine; analgesic; haemostatic; thrombolytic; antiinflammatory;  
CC cytostatic; antibacterial; antifungal; antiviral; antidiabetic;  
CC anasthmatic; vulnary; antitumor; osteoprotective; neuroprotective;  
CC nootropic; antiparkinsonian; anticonvulsant; cerebroprotective;  
CC anticonvulsant; and antidepressant. The SESTs can be used for gene  
CC therapy and in vaccines. The SESTs are useful as probes for the  
CC identification and isolation of full-length cDNAs and genomic DNA  
CC molecules which correspond to the SESTs. Proteins encoded by the SESTs  
CC are useful in assays for determining biological activity and raising  
CC antibodies. They may be useful for treatment of autoimmune disorders  
CC (multiple sclerosis, insulin dependent diabetes), allergic conditions  
CC (asthma), myeloid or lymphoid cell deficiencies, wounds, burns, ulcers,  
CC osteoporosis, osteoarthritis, central nervous system disorders  
CC (Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation  
CC disorders (haemophilia, thrombosis), inflammatory disorders (Crohn's  
CC disease), tumours, bacterial, fungal or viral infections, depression and  
CC psoriasis. AAA45926 to AAA45931 represent linker variants which are given  
XX in the exemplification of the present invention.  
XX  
XX Sequence 450 BP; 134 A; 81 C; 76 G; 159 T; 0 other;

Query Match 3.9%; Score 28; DB 21; Length 450;  
Best Local Similarity 100.0%; Pred. No. 0.052;  
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 449 TTTTTTTTTTTTTTTTTTTTTGGCTTC 476  
|||||  
Db 309 ttttttttttttttttttttttttggcttc 336

RESULT 8  
AAZ97350/C  
ID AAZ97350 standard; cDNA; 499 BP.  
XX  
XX AAZ97350:  
XX  
XX 18-APR-2000 (first entry)  
XX Human prostate cancer differentially expressed gene #211.  
XX  
XX Prostate cancer specific gene; cancer; tumour progression; diagnose;  
XX hyperproliferative cell growth; prostatic disorder; treatment;  
XX metastatic prostate cancer; benign prostate hyperplasia; BPH; ss.  
XX  
XX Homo sapiens.  
XX  
XX WO9964594-A2.  
XX  
XX 16-DEC-1999.  
XX  
XX 10-JUN-1999; 99WO-US13181.



CC	isolated from a range of human tissues disclosed in the specification.
CC	The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC	in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC	and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC	marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC	(b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC	hemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC	disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC	colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC	(d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC	epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC	and parasitic infections.
CC	Note: The sequence data for this patent did not form part of the
CC	printed specification, but was obtained in electronic format directly
CC	from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX	
SQ	Sequence 626 BP: 103 A; 181 C; 141 G; 201 T; 0 other:
	Query Match            3.9%   Score 28; DB 22; Length 626;
	Best Local Similarity 100.0%; Pred. No. 0.049;
	Matches     28; Conservative     0; Mismatches     0; Indels     0; Gaps     0;
QY	443 CCACCTCTTTTTTTTTTTTTTTTTTT 470
DB	1 ccaactttttttttttttttttttttt 28
RESULT 10	
ID	ABR21398 standard; DNA; 626 BP.
AC	ABA21398;
DT	23-JAN-2002 (first entry)
DE	Human nervous system related polynucleotide SEQ ID NO 13729.
KW	Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW	immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW	antiParkinsonian; antistickling; antianaemic; antiarthritis; cancer;
KW	antithumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW	antiallergic; antidiabetic; antitumor; anticovulsant; antifungal;
KW	antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW	neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
OS	Homo sapiens.
PX	WO2001159063-A2.
PN	16-AUG-2001.
PD	
PP	17-JAN-2001; 2001WO-US01334.
PR	31-JAN-2000; 2000US-0179065
PR	04-FEB-2000; 2000US-0180628
PR	24-FEB-2000; 2000US-0184564
PR	02-MAR-2000; 2000US-0186350
PR	16-MAR-2000; 2000US-0189874
PR	17-MAR-2000; 2000US-0190076
PR	18-APR-2000; 2000US-0198123
PR	19-MAY-2000; 2000US-0205515
PR	07-JUN-2000; 2000US-0209467
PR	28-JUN-2000; 2000US-0214886
PR	30-JUN-2000; 2000US-0215135
PR	07-JUL-2000; 2000US-0216647
PR	07-JUL-2000; 2000US-0216880
PR	11-JUL-2000; 2000US-0217487
PR	11-JUL-2000; 2000US-0217496
PR	14-JUL-2000; 2000US-0218290
PR	26-JUL-2000; 2000US-0220963
PR	26-JUL-2000; 2000US-0220964
PR	14-AUG-2000; 2000US-0224518



CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 626 BP; 103 A; 181 C; 141 G; 201 T; 0 other;

Query Match 3.9%; Score 28; DB 22; Length 626;  
Best Local Similarity 100.0%; Pred. No. 0.049;  
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 443 CCACCTCTTTTTTTTTTTTTTTTTTTTTTTT 470  
|||||

Db 1 ccactctctctctctctctctctctctctctt 28  
|||||

RESULT 11

AA11085/c

ID AAL11085 standard; cDNA; 761 BP.

AC AAL11085;

DT 07-DEC-2001 (first entry);

DE Human breast cancer expressed polynucleotide 3542.

XX Human; breast cancer; cell marker; cytostatic; ss.

XX Homo sapiens.

XX WO200151628-A2.

XX 19-JUL-2001.

XX 10-JAN-2001; 2001WO-US00798.

XX 14-JAN-2000; 2000US-0176077.

XX 14-MAR-2000; 2000US-0189167.

XX 24-MAR-2000; 2000US-0192099.

XX 29-MAR-2000; 2000US-0193480.

XX 15-MAY-2000; 2000US-0205230.

XX 09-JUN-2000; 2000US-0211315.

XX 25-JUL-2000; 2000US-0220534.

XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX Lillie J, Xu Y, Wang Y, Steinmann K;

XX WPI; 2001-451856/48.

XX New peptide useful as a marker for the diagnosis of breast cancer -

XX Claim 1; Page 652-653; 3695pp; English.

XX The invention relates to human breast cancer expressed polynucleotides

XX (AA107544-AA126789) and methods of assessing whether a patient is

XX afflicted with breast cancer by examining the correlation between the

XX expression of certain markers and the cancerous state of breast cells.

XX The polynucleotides and encoded polypeptides are potential markers for

XX detecting, diagnosing, monitoring, characterizing treating and

XX potentially preventing breast cancer. The polynucleotides and encoded

XX polypeptides are also useful for isolating compounds with cytostatic

XX activity.

XX Sequence 761 BP; 183 A; 186 C; 173 G; 192 T; 27 other;

Query Match 3.9%; Score 28; DB 22; Length 761;

Best Local Similarity 100.0%; Pred. No. 0.048;

Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 449 TTTTCTTTTTTTTTTTTTTTTTTTTGGCTTC 476

|||||

Db 691 TTTTCTTTTTTTTTTTTTTTTTTTTGGCTTC 664

RESULT 12

ABA02499/c

ID ABA02499 standard; cDNA; 1008 BP.

XX ABA02499;

XX 25-MAR-2002 (first entry)

XX Maize pathogen resistance-associated protein Rarl1-encoding cDNA.

XX Maize; Rarl1; pathogen resistance-associated; plant; HR;

XX hypersensitivity response; viral infection; fungal infection;

XX transgenic; gene; ss.

XX Zea mays.

XX Location/Qualifiers

XX 1..34

XX /tag= a

XX 35..712

XX /tag= b

XX /product= "Maize Rarl1 protein"

XX 713..929

XX /tag= c

XX US2001047521-A1.

XX 29-NOV-2001.

XX 09-FEB-2001; 2001US-0780641.

XX 09-FEB-2000; 2000US-181256P.

XX (CRAN/) CRANE E H.

XX (PIFF/) PIFFANELLI P.

XX (SIMM/) SIMMONS C R.

XX Crane EH, Piffanelli P, Simmons CR;

XX WPI: 2002-105616/14.

XX P-PSDB; AAM53060.

XX New polynucleotides, useful for improving the virus-resistance of

XX plants such as maize, soybean, sunflower, sorghum, canola, wheat,

XX alfalfa, cotton, rice, barley, and millet, comprise nucleic acids

XX encoding a Zea mays Rarl1 proteins -

XX Claim 1; Page 21-22; 24pp; English.

XX This sequence represents cDNA encoding the maize pathogen resistance-

XX associated protein Rarl1. The maize Rarl1 protein has 79% homology to the

XX barley Rarl1 protein over its entire length. The Rarl1 protein is required

XX for the function of certain plant resistance (R) gene products which,

XX along with pathogen avirulence (avr) gene products, mediate the

XX hypersensitivity response (HR), a series of physical and physiological

XX changes in host cells local to the site of infection which limits the

XX spread of the infection. Rarl1 nucleic acids may be used in the generation

XX of transgenic plants and seeds that exhibit increased expression of

XX the Rarl1 protein. They may be used in this way to enhance resistance to

XX a variety of pathogens, particularly viruses and fungi (e.g., powdery

XX mildew) in plants such as maize, soybean, sunflower, sorghum, canola,

XX wheat, alfalfa, cotton, rice, barley, and millet.

XX Sequence 1008 BP; 331 A; 215 C; 243 G; 219 T; 0 other;

Query Match 3.9%; Score 28; DB 24; Length 1008;

Best Local Similarity 100.0%; Pred. No. 0.046;

Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 449 TTTTCTTTTTTTTTTTTTTTTTTTTGGCTTC 476



PN	W0200157182-A2.
XX	
PD	
XX	09-AUG-2001.
XX	
PF	17-JAN-2001; 2001WO-US01354.
XX	
XX	31-JAN-2000; 2000US-0179065.
PR	04-FEB-2000; 2000US-0180628.
PR	24-FEB-2000; 2000US-0184664.
PR	02-MAR-2000; 2000US-0186350.
PR	16-MAR-2000; 2000US-0189874.
PR	17-MAR-2000; 2000US-0190076.
PR	18-APR-2000; 2000US-0198123.
PR	19-MAY-2000; 2000US-0205515.
PR	07-JUN-2000; 2000US-0209467.
PR	28-JUN-2000; 2000US-0211385.
PR	30-JUN-2000; 2000US-0215135.
PR	07-JUL-2000; 2000US-0216647.
PR	07-JUL-2000; 2000US-0216880.
PR	11-JUL-2000; 2000US-0217487.
PR	11-JUL-2000; 2000US-0218290.
PR	14-JUL-2000; 2000US-0220963.
PR	26-JUL-2000; 2000US-0220964.
PR	16-AUG-2000; 2000US-0224518.
PR	14-AUG-2000; 2000US-0224519.
PR	14-AUG-2000; 2000US-0225213.
PR	14-AUG-2000; 2000US-0225214.
PR	14-AUG-2000; 2000US-0225266.
PR	14-AUG-2000; 2000US-0225267.
PR	14-AUG-2000; 2000US-0226270.
PR	14-AUG-2000; 2000US-0226447.
PR	14-AUG-2000; 2000US-0227557.
PR	14-AUG-2000; 2000US-0227558.
PR	18-AUG-2000; 2000US-0227579.
PR	22-AUG-2000; 2000US-0228621.
PR	22-AUG-2000; 2000US-0228668.
PR	22-AUG-2000; 2000US-0228668.
PR	22-AUG-2000; 2000US-0327182.
PR	23-AUG-2000; 2000US-03277009.
PR	30-AUG-2000; 2000US-0328924.
PR	01-SEP-2000; 2000US-0329287.
PR	01-SEP-2000; 2000US-0329343.
PR	01-SEP-2000; 2000US-0329344.
PR	01-SEP-2000; 2000US-0329345.
PR	05-SEP-2000; 2000US-0329509.
PR	05-SEP-2000; 2000US-0329513.
PR	06-SEP-2000; 2000US-03230437.
PR	06-SEP-2000; 2000US-03230438.
PR	08-SEP-2000; 2000US-03231242.
PR	08-SEP-2000; 2000US-03231243.
PR	08-SEP-2000; 2000US-03231244.
PR	08-SEP-2000; 2000US-03231414.
PR	08-SEP-2000; 2000US-03232080.
PR	12-SEP-2000; 2000US-0323081.
PR	12-SEP-2000; 2000US-0321968.
PR	14-SEP-2000; 2000US-0323397.
PR	14-SEP-2000; 2000US-0323398.
PR	14-SEP-2000; 2000US-0323399.
PR	14-SEP-2000; 2000US-0323401.
PR	14-SEP-2000; 2000US-0323063.
PR	14-SEP-2000; 2000US-0323064.
PR	21-SEP-2000; 2000US-0323065.
PR	21-SEP-2000; 2000US-03234223.
PR	21-SEP-2000; 2000US-03234274.
PR	25-SEP-2000; 2000US-0323497.
PR	25-SEP-2000; 2000US-0323498.
PR	26-SEP-2000; 2000US-03234984.
PR	27-SEP-2000; 2000US-03235834.
PR	27-SEP-2000; 2000US-03235836.

PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249246.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
(HUMA-) HUMAN GENOME SCI INC.  
Rosen CA, Barash SC, Ruben SM;

XX WPI; 2001-483426/52.  
DR Nucleic acids encoding human immune/haematopoietic antigen polypeptides,  
XX useful for preventing, diagnosing and/or treating cancers and  
PT metastasis -  
PT  
XX  
PS Disclosure; SEQ ID NO 25700; 3071pp + Sequence Listing; English.  
XX  
XX AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)  
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic  
CC activity, and can be used in gene therapy and vaccine production. (I)  
CC proteins, and polynucleotides may be used in the prevention, diagnosis and  
CC treatment of diseases associated with inappropriate (I) expression. For  
CC example, they may be used to treat disorders associated with decreased  
CC expression by rectifying mutations or deletions in a patient's genome  
CC that affect the activity of (I) by expressing inactive proteins or to  
CC supplement the patients own production of (I). Additionally, (I)  
CC polynucleotides may be used to produce the secreted (I), by inserting the  
CC the nucleic acids into a host cell and culturing the cell to express the  
CC protein. (I) proteins and polynucleotides may be used to prevent,  
CC diagnose and treat immune/haematopoietic-related diseases, especially  
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703  
CC to AAK87694 represent human immune/haematopoietic antigen genomic  
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169  
CC represent sequences used in the exemplification of the present invention.  
XX  
SQ Sequence 1086 BP; 312 A; 302 C; 233 G; 239 T; 0 other;  
  
Query Match 3.9%; Score 28; DB 22; Length 1086;  
Best Local Similarity 100.0%; Pred. No. 0.045;  
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 443 CCACCTCTTTTCTTTTCTTTTCTTTTCTTTT 470  
Db 5 ccactctttttttttttttttttttttttttttt 32  
  
RESULT 15  
AAK84102  
ID AAK84102 standard; DNA; 1086 BP.  
XX  
AC AAK84102;  
XX  
DT 07-NOV-2001 (first entry)  
XX  
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:38914.  
XX  
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;  
KW cytostatic; gene therapy; vaccine; metastasis; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200157182-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US01354.  
XX  
PR 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.



PR 11-JUL-2000; 2000US-0217487.  
 PR 11-JUL-2000; 2000US-0217496.  
 PR 14-JUL-2000; 2000US-0218290.  
 PR 26-JUL-2000; 2000US-0218296.  
 PR 26-JUL-2000; 2000US-0220963.  
 PR 14-AUG-2000; 2000US-0220964.  
 PR 14-AUG-2000; 2000US-0224518.  
 PR 14-AUG-2000; 2000US-0224519.  
 PR 14-AUG-2000; 2000US-0225213.  
 PR 14-AUG-2000; 2000US-0225214.  
 PR 14-AUG-2000; 2000US-0225266.  
 PR 14-AUG-2000; 2000US-0225267.  
 PR 14-AUG-2000; 2000US-0225268.  
 PR 14-AUG-2000; 2000US-0225270.  
 PR 14-AUG-2000; 2000US-0225447.  
 PR 14-AUG-2000; 2000US-0225447.  
 PR 14-AUG-2000; 2000US-0225757.  
 PR 14-AUG-2000; 2000US-0225758.  
 PR 14-AUG-2000; 2000US-0225759.  
 PR 18-AUG-2000; 2000US-0226279.  
 PR 22-AUG-2000; 2000US-0226681.  
 PR 22-AUG-2000; 2000US-0226686.  
 PR 22-AUG-2000; 2000US-0227182.  
 PR 23-AUG-2000; 2000US-0227009.  
 PR 30-AUG-2000; 2000US-0228924.  
 PR 01-SEP-2000; 2000US-0229287.  
 PR 01-SEP-2000; 2000US-0229343.  
 PR 01-SEP-2000; 2000US-0229344.  
 PR 01-SEP-2000; 2000US-0229345.  
 PR 03-SEP-2000; 2000US-0229509.  
 PR 03-SEP-2000; 2000US-0229513.  
 PR 06-SEP-2000; 2000US-0230437.  
 PR 06-SEP-2000; 2000US-0230438.  
 PR 08-SEP-2000; 2000US-0231242.  
 PR 08-SEP-2000; 2000US-0231243.  
 PR 08-SEP-2000; 2000US-0231244.  
 PR 08-SEP-2000; 2000US-0231413.  
 PR 08-SEP-2000; 2000US-0231414.  
 PR 08-SEP-2000; 2000US-0232080.  
 PR 08-SEP-2000; 2000US-0232081.  
 PR 12-SEP-2000; 2000US-0231968.  
 PR 14-SEP-2000; 2000US-0232397.  
 PR 14-SEP-2000; 2000US-0232398.  
 PR 14-SEP-2000; 2000US-0232399.  
 PR 14-SEP-2000; 2000US-0232400.  
 PR 14-SEP-2000; 2000US-0232401.  
 PR 14-SEP-2000; 2000US-0233063.  
 PR 14-SEP-2000; 2000US-0233064.  
 PR 14-SEP-2000; 2000US-0233065.  
 PR 21-SEP-2000; 2000US-0234223.  
 PR 21-SEP-2000; 2000US-0234223.  
 PR 25-SEP-2000; 2000US-0234697.  
 PR 25-SEP-2000; 2000US-0234697.  
 PR 26-SEP-2000; 2000US-0234696.  
 PR 26-SEP-2000; 2000US-0234696.  
 PR 27-SEP-2000; 2000US-0235834.  
 PR 27-SEP-2000; 2000US-0235836.  
 PR 29-SEP-2000; 2000US-0236327.  
 PR 29-SEP-2000; 2000US-0236367.  
 PR 29-SEP-2000; 2000US-0236368.  
 PR 29-SEP-2000; 2000US-0236369.  
 PR 29-SEP-2000; 2000US-0236370.  
 PR 02-OCT-2000; 2000US-0236802.  
 PR 02-OCT-2000; 2000US-0237037.  
 PR 02-OCT-2000; 2000US-0237038.  
 PR 02-OCT-2000; 2000US-0237039.  
 PR 02-OCT-2000; 2000US-0237040.  
 PR 13-OCT-2000; 2000US-0239935.  
 PR 13-OCT-2000; 2000US-0239937.  
 PR 20-OCT-2000; 2000US-0240960.  
 PR 20-OCT-2000; 2000US-0241221.  
 PR 20-OCT-2000; 2000US-0241785.  
 PR 20-OCT-2000; 2000US-0241785.  
 PR 20-OCT-2000; 2000US-0241787.  
 PR 20-OCT-2000; 2000US-0241808.  
 PR 20-OCT-2000; 2000US-0241809.

PR 20-OCT-2000; 2000US-0241826.  
 PR 01-NOV-2000; 2000US-0244617.  
 PR 08-NOV-2000; 2000US-0246474.  
 PR 08-NOV-2000; 2000US-0246475.  
 PR 08-NOV-2000; 2000US-0246476.  
 PR 08-NOV-2000; 2000US-0246477.  
 PR 08-NOV-2000; 2000US-0246478.  
 PR 08-NOV-2000; 2000US-0246523.  
 PR 08-NOV-2000; 2000US-0246524.  
 PR 08-NOV-2000; 2000US-0246525.  
 PR 08-NOV-2000; 2000US-0246526.  
 PR 08-NOV-2000; 2000US-0246527.  
 PR 08-NOV-2000; 2000US-0246528.  
 PR 08-NOV-2000; 2000US-0246532.  
 PR 08-NOV-2000; 2000US-0246609.  
 PR 08-NOV-2000; 2000US-0246610.  
 PR 08-NOV-2000; 2000US-0246611.  
 PR 08-NOV-2000; 2000US-0246613.  
 PR 17-NOV-2000; 2000US-0249207.  
 PR 17-NOV-2000; 2000US-0249208.  
 PR 17-NOV-2000; 2000US-0249209.  
 PR 17-NOV-2000; 2000US-0249210.  
 PR 17-NOV-2000; 2000US-0249211.  
 PR 17-NOV-2000; 2000US-0249212.  
 PR 17-NOV-2000; 2000US-0249213.  
 PR 17-NOV-2000; 2000US-0249214.  
 PR 17-NOV-2000; 2000US-0249215.  
 PR 17-NOV-2000; 2000US-0249216.  
 PR 17-NOV-2000; 2000US-0249217.  
 PR 17-NOV-2000; 2000US-0249218.  
 PR 17-NOV-2000; 2000US-0249244.  
 PR 17-NOV-2000; 2000US-0249245.  
 PR 17-NOV-2000; 2000US-0249264.  
 PR 17-NOV-2000; 2000US-0249265.  
 PR 17-NOV-2000; 2000US-0249297.  
 PR 17-NOV-2000; 2000US-0249299.  
 PR 17-NOV-2000; 2000US-0249300.  
 PR 01-DEC-2000; 2000US-0250160.  
 PR 01-DEC-2000; 2000US-0250391.  
 PR 05-DEC-2000; 2000US-0251030.  
 PR 05-DEC-2000; 2000US-0251988.  
 PR 05-DEC-2000; 2000US-0256719.  
 PR 06-DEC-2000; 2000US-0251479.  
 PR 08-DEC-2000; 2000US-0251856.  
 PR 08-DEC-2000; 2000US-0251868.  
 PR 08-DEC-2000; 2000US-0251869.  
 PR 08-DEC-2000; 2000US-0251989.  
 PR 08-DEC-2000; 2000US-0251990.  
 PR 11-DEC-2000; 2000US-0254097.  
 PR 05-JAN-2001; 2001US-0259678.  
 XX  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 PA  
 XX Rosen CA, Barash SC, Ruben SM;  
 PI  
 XX WPI; 2001-483426/52.  
 XX  
 XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
 PT useful for preventing, diagnosing and/or treating cancers and  
 PT metastasis -  
 XX  
 XX Disclosure; SEQ ID NO 38914; 3071pp + Sequence Listing; English.  
 XX  
 XX AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)  
 CC amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic  
 CC activity, and can be used in gene therapy and vaccine production. (I)  
 CC proteins and polynucleotides may be used in the prevention, diagnosis and  
 CC treatment of diseases associated with inappropriate (I) expression. For  
 CC example, they may be used to treat disorders associated with decreased  
 CC expression by rectifying mutations or deletions in a patient's genome  
 CC that affect the activity of (I) by expressing inactive proteins or to  
 CC supplement the patient's own production of (I). Additionally, (I)  
 CC polynucleotides may be used to produce the secreted (I), by inserting

XX SQ

Query Match 3.9%; Score 28; DB 22; Length 1086;  
Best Local Similarity 100.0%; Pred. No. 0.045;  
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

D**b**

Search completed: August 3, 2002, 01:50:00  
Job time: 32185 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 2, 2002, 14:19:05 ; Search time 7016.61 Seconds  
(without alignments)  
1390.743 Million cell updates/sec

Title: US-08-973-363-1

Perfect score: 723

Sequence: 1. CCCGGTCGAGGTTTCAAGG.....CTTGAAACCTCCGACCGGG 723

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 13736207 seqs, 6748477542 residues

Word size : 0

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:\*

- 1: em\_estba:\*
- 2: em\_esthum:\*
- 3: em\_estlin:\*
- 4: em\_estmu:\*
- 5: em\_estov:\*
- 6: em\_estpi:\*
- 7: em\_estro:\*
- 8: em\_hic:\*
- 9: gb\_est1:\*
- 10: gb\_est2:\*
- 11: gb\_hic:\*
- 12: gb\_gss:\*
- 13: em\_gss\_hum:\*
- 14: em\_gss\_inv:\*
- 15: em\_gss\_pln:\*
- 16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	31	4.3	382	9	AI630307
2	31	4.3	130	11	BC022284
3	30	4.1	313	9	AI916488
4	30	4.1	356	12	A2802813
5	30	4.1	359	9	AI607901
6	30	4.1	361	12	A2064574
7	30	4.1	364	10	BI739401
8	30	4.1	388	12	AQ08063
9	30	4.1	501	9	AI744071
10	30	4.1	516	10	BE652261
11	30	4.1	535	10	BI404399
12	30	4.1	544	12	BH317627
13	30	4.1	560	12	BH314151
14	30	4.1	573	12	BH316471
15	30	4.1	585	9	AA946504
16	30	4.1	611	10	BF472409
17	30	4.1	647	12	AG144824

c	18	30	4.1	662	10	BF673589
	19	30	4.1	811	12	AQ752424
	20	29	4.0	111	9	AI362328
	21	29	4.0	147	12	A2250555
	22	29	4.0	164	9	AI511169
c	23	29	4.0	221	9	AW301935
	24	29	4.0	253	12	AZ494100
	25	29	4.0	260	10	BI865453
c	26	29	4.0	285	10	BF415725
	27	29	4.0	294	9	AI566778
	28	29	4.0	329	9	AI770139
	29	29	4.0	354	12	AQ538422
	30	29	4.0	399	9	AI702770
	31	29	4.0	401	9	AI206568
	32	29	4.0	422	9	AI954748
	33	29	4.0	426	9	AA153854
	34	29	4.0	427	9	AA140515
	35	29	4.0	462	12	AZ057128
	36	29	4.0	483	9	AI193285
	37	29	4.0	489	9	AA713517
c	38	29	4.0	551	12	AZ589435
	39	29	4.0	552	10	BG453842
	40	29	4.0	558	9	AL511215
c	41	29	4.0	563	12	AZ362131
c	42	29	4.0	566	12	BH089139
	43	29	4.0	573	12	AZ101654
	44	29	4.0	575	9	AL511339
	45	29	4.0	586	12	BH305532

#### ALIGNMENTS

#### RESULT 1

AI630307  
LOCUS ad08d01.y1 Proliferating Erythroid Cells (LCB:ad library) Homo sapiens CDNA clone ad08d01 random, mRNA sequence.  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
COMMENT

AI630307 382 bp mRNA linear EST 08-MAR-2000  
ad08d01.y1 Proliferating Erythroid Cells (LCB:ad library) Homo sapiens CDNA clone ad08d01 random, mRNA sequence.  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
COMMENT

#### FEATURES

source  
Location/Qualifiers  
1. 382  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="ad08d01"  
/clone\_lib="Proliferating Erythroid Cells (LCB:ad library)"  
/sex="unknown"

```
/tissue_type="blood"
/cell_type="Erythroid Cells"
/lab_host="Primary Culture of Peripheral Blood
Mononuclear Cells"
/dev_stage="Progenitor; EPO responsive CD71++++"
/lab_host="DH5alpha"
/Note="Human blood; Vector: pCRII; Site_1: EcoRI; Site_2:
EcoRI; Human peripheral blood mononuclear cells were
cultured in the presence or absence of erythropoietin.
CD71+++ cells arising only in erythropoietin-supplemented
medium were isolated by fluorescence activated cell
sorting. Those cells demonstrated an average of six
additional doublings in suspension culture and erythroid
colony formation in methylcellulose. Suppression
subtractive hybridization was used to construct the ad
library (tester-sorted CD71+++ cells, driver=unsorted
cells cultured without erythropoietin)."
BASE COUNT      82 a      87 c      60 g      153 t
ORIGIN

Query Match      4.3%; Score 31; DB 9; Length 382;
Best Local Similarity 100.0%; Pred. No. 3.7e+02;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 449 TTTTGTGCTCTTC 479
|||||
Db 4 TTTTGTGCTCTTC 34

RESULT 2
BC022284/c
LOCUS      1130 bp      mRNA      linear      HTC 04-FEB-2002
DEFINITION Homo sapiens, clone IMAGE:4762273, mRNA.
ACCESSION  BC022284
VERSION     BC022284.1 GI:18490147
KEYWORDS   HTC.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            1 (bases 1 to 1130)
Strausberg,R.
Direct Submission
Submitted (01-FEB-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgapbs@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: CLONTECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: http://www-shgc.stanford.edu
Contact: (Dickson, Mark) mcdpaxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
R. M.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAL Plate: 37 Row: j Column: 5
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 4885172
This clone has the following problem: frame shifted.
Location/Qualifiers
1..1130
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4762273"
/tissue_type="Lung, mucoepidermoid carcinoma"

FEATURES
source
```

```
/clone_lib="NIH_MGC_59"
/lab_host="DH10B"
/Note="Vector: pDNR-LIB"
BASE COUNT      346 a      277 c      319 g      188 t
ORIGIN

Query Match      4.3%; Score 31; DB 11; Length 1130;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 449 TTTTGTGCTCTTC 479
|||||
Db 1122 TTTTGTGCTCTTC 1092

RESULT 3
AI916488      313 bp      mRNA      linear      EST 16-DEC-1999
LOCUS      tz71d01.x1 NCI-CGAP_Pan1 Homo sapiens cDNA clone IMAGE:2294017 3',
DEFINITION mRNA sequence.
ACCESSION  AI916488
VERSION     AI916488.1 GI:5636343
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            1 (bases 1 to 313)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Life Technologies catalog #: 11548-013
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/dbp/image/image.html
Insert Length: 2643 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 222.
Location/Qualifiers
1..313
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2294017"
/clone_lib="NCI-CGAP_Pan1"
/tissue_type="adenocarcinoma"
/lab_host="DH10B"
/Note="Organ: pancreas; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.72 kb. Life Technologies catalog #:
11548-013"
BASE COUNT      51 a      82 c      91 g      89 t
ORIGIN

Query Match      4.1%; Score 30; DB 9; Length 313;
Best Local Similarity 100.0%; Pred. No. 7.5e+02;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 449 TTTTGTGCTCTTC 478
|||||
Db 9 TTTTGTGCTCTTC 38

RESULT 4
AZ802813/c
LOCUS      356 bp      DNA      linear      GSS 16-FEB-2001
DEFINITION 2M0061B20R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
clone UUGC2M0061B20 R, DNA sequence.
```





Matches	30;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
QY	449	TTTTTTTTTTTTTTTTTTTTTTTTTGCTCTT	478						
DB	40	TTTTTTTTTTTTTTTTTTTTTTTTTGCTCTT	69						
RESULT	9								
LOCUS	AI744071								
DEFINITION	AI744071	501 bp	mrna	linear	EST 17-DEC-1999				
REFERENCE	AI744071	501 bp	mrna	linear	EST 17-DEC-1999				
AUTHORS	AI744071	501 bp	mrna	linear	EST 17-DEC-1999				
TITLE	AI744071	501 bp	mrna	linear	EST 17-DEC-1999				
JOURNAL	AI744071	501 bp	mrna	linear	EST 17-DEC-1999				
COMMENT	AI744071	501 bp	mrna	linear	EST 17-DEC-1999				
ACCESSION	AI744071	501 bp	mrna	linear	EST 17-DEC-1999				
VERSION	AI744071	501 bp	mrna	linear	EST 17-DEC-1999				
KEYWORDS	AI744071	501 bp	mrna	linear	EST 17-DEC-1999				
SOURCE	AI744071	501 bp	mrna	linear	EST 17-DEC-1999				
ORGANISM	AI744071	501 bp	mrna	linear	EST 17-DEC-1999				
REFERENCE	AI744071	501 bp	mrna	linear	EST 17-DEC-1999				
AUTHORS	AI744071	501 bp	mrna	linear	EST 17-DEC-1999				
TITLE	AI744071	501 bp	mrna	linear	EST 17-DEC-1999				
JOURNAL	AI744071	501 bp	mrna	linear	EST 17-DEC-1999				
COMMENT	AI744071	501 bp	mrna	linear	EST 17-DEC-1999				
ACCESSION	AI744071	501 bp	mrna	linear	EST 17-DEC-1999				
VERSION	AI744071	501 bp	mrna	linear	EST 17-DEC-1999				
KEYWORDS	AI744071	501 bp	mrna	linear	EST 17-DEC-1999				
SOURCE	AI744071	501 bp	mrna	linear	EST 17-DEC-1999				
ORGANISM	AI744071	501 bp	mrna	linear	EST 17-DEC-1999				
REFERENCE	AI744071	501 bp	mrna	linear	EST 17-DEC-1999				
AUTHORS	AI744071	501 bp	mrna	linear	EST 17-DEC-1999				
TITLE	AI744071	501 bp	mrna	linear	EST 17-DEC-1999				
JOURNAL	AI744071	501 bp	mrna	linear	EST 17-DEC-1999				
COMMENT	AI744071	501 bp	mrna	linear	EST 17-DEC-1999				
ACCESSION	AI744071	501 bp	mrna	linear	EST 17-DEC-1999				
VERSION	AI744071	501 bp	mrna	linear	EST 17-DEC-1999				
KEYWORDS	AI744071	501 bp	mrna	linear	EST 17-DEC-1999				
SOURCE	AI744071	501 bp	mrna	linear	EST 17-DEC-1999				
ORGANISM	AI744071	501 bp	mrna	linear	EST 17-DEC-1999				
REFERENCE	AI744071	501 bp	mrna	linear	EST 17-DEC-1999				
AUTHORS	AI744071	501 bp	mrna	linear	EST 17-DEC-1999				
TITLE	AI744071	501 bp	mrna	linear	EST 17-DEC-1999				
JOURNAL	AI744071	501 bp	mrna	linear	EST 17-DEC-1999				
COMMENT	AI744071	501 bp	mrna	linear	EST 17-DEC-1999				
ACCESSION	AI744071	501 bp	mrna	linear	EST 17-DEC-1999				
VERSION	AI744071	501 bp	mrna	linear	EST 17-DEC-1999				
KEYWORDS	AI744071	501 bp	mrna	linear	EST 17-DEC-1999				
SOURCE	AI744071	501 bp	mrna	linear	EST 17-DEC-1999				
ORGANISM	AI744071	501 bp	mrna	linear	EST 17-DEC-1999				
REFERENCE	AI744071	501 bp	mrna						

**AUTHORS** Bonaldo,M.F., Lennon,G. and Soares,M.B.  
**TITLE** Normalization and subtraction: two approaches to facilitate gene discovery  
**JOURNAL** Genome Res. 6 (9), 791-806 (1996)  
**MEDLINE** 9704477  
**COMMENT** Contact: Tuggie CK  
Molecular Genetics Laboratory, Department of Animal Science  
Iowa State University  
201 Kildee Hall, Ames, IA 50011-3150, USA  
Tel: 5152944252  
Fax: 5152942401  
Email: cktuggie@iastate.edu  
The sequence contained an oligo-dT track that was present in the strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to verify it as a clone from the normalized uterus library cDNA Library Preparation: M.B. Soares Lab , University of Iowa EST sequencing: M.B. Soares Lab, University of Iowa Clone distribution: clones will be available through Research Genetics (www.resgen.com)  
Seq primer: M13 Forward  
FOLIA-fes.

**FEATURES**

source

Location/Qualifiers

1. .535

/organism="Sus scrofa"

/strain="crossbreed"

/db\_xref="taxon:9823"

/clone="MI-P-CP1-nwz-b-06-0-UI"

/lab\_host="MI-P-CP1"

/note="Vector: pT73D-Pac (Pharmacia)"

polylinker: Site 1: Not I; Site 2: EcoRI; The MI-P-CP1 library is normalized library derived from the MT-P-CP0 library, ultimately derived from uterus tissue. For a detailed description of the library from which this clone was derived, please visit our web site at <http://pigest.genome.iastate.edu/>. The procedure used to create this library has been previously described (Bonaldo , Lennon and Soares, Genome Research 6: 791-806, 1996)

TAG\_LIB="MI-P-CP1"

TAG\_TISSUE="uterus"

TAG\_SEQ="AGTCCATCG"

149 a 126 c 83 g 177 t

BASE COUNT  
ORIGIN

Query Match 4.1%; Score 30; DB 10; Length 535;  
Best Local Similarity 100.0%; Pred. No. 4.3e+02;  
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 453 TTTTGTGCTCTTCATA 482

|||||

Db 1 TTTTGTGCTCTTCATA 30

**RESULT**

12

BH317627

LOCUS

BH317627 544 bp DNA linear GSS 03-DEC-2001

CH230-159G24.TJ CHORI-230 Segment 1 Rattus norvegicus genomic clone

CH230-159G24, DNA sequence.

ACCESSION

BH317627

VERSION

BH317627.1

KEYWORDS

GSS

SOURCE

Norway rat.

ORGANISM

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

REFERENCE

1 (bases 1 to 544)

AUTHORS

Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K., Shvartsbeyn

,A., Gebregeorgis,E., Overton,L., Russell,D., Chen,D., Riggs,F., de

**TITLE** Rat BAC End Sequences from Library CHORI-230 EcoRI segment  
**JOURNAL** Unpublished (1999)  
**COMMENT** Other\_GSSs: CH230-159G24.TV  
Contact: Shaying Zhao  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: szhao@tigr.org

Clones are derived from the rat BAC library CHORI-230  
(<http://www.chori.org/bacpac/rat230.htm>). For BAC library  
availability, please contact Pieter de Jong ([pdejong@mail.cho.org](mailto:pdejong@mail.cho.org)).  
Clones may be purchased from BACPAC Resources  
(<http://www.chori.org/bacpac/orering.information.htm>). BAC end  
page: [http://www.tigr.org/tdb/bac\\_ends/rat/bac\\_end\\_intro.html](http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html)  
Plate: 159 row: G column: 24  
Seq primer: SP6  
Class: BAC ends.

**FEATURES** Location/Qualifiers

1. .544

/organism="Rattus norvegicus"

/strain="BN/SSNhsd/MCW"

/db\_xref="taxon:10116"

/clone="CH230-159G24"

/clone\_lib="CHORI-230 Segment 1"

/sex="Female"

/cell\_type="Brain"

/note="Vector: PTARBAC2.1; Site\_1: EcoRI; Site\_2: EcoRI;

CHORI-230 Rat (BN/SSNhsd/MCW) BAC library produced by

Pieter de Jong"

BASE COUNT 97 a 188 c 119 g 140 t

ORIGIN

Query Match 4.1%; Score 30; DB 12; Length 544;

Best Local Similarity 100.0%; Pred. No. 4.3e+02;

Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 443 CCACCTCTTTTGTGCTCTTCATA 472

|||||

Db 311 CCACCTCTTTTGTGCTCTTCATA 340

RESULT 13

BH314151

LOCUS

BH314151 560 bp DNA linear GSS 03-DEC-2001

CH230-159A24.TJB CHORI-230 Segment 1 Rattus norvegicus genomic

clone CH230-159A24, DNA sequence.

ACCESSION

BH314151

VERSION

BH314151.1

KEYWORDS

GSS

SOURCE

Norway rat.

ORGANISM

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

REFERENCE

1 (bases 1 to 560)

AUTHORS

Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K., Shvartsbeyn

,A., Gebregeorgis,E., Overton,L., Russell,D., Chen,D., Riggs,F., de

Jong,P. and Fraser,C.M.

Rat BAC End Sequences from Library CHORI-230 EcoRI segment

Unpublished (1999)

Other\_GSSs: CH230-159A24.TV

Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: szhao@tigr.org

Clones are derived from the rat BAC library CHORI-230

(<http://www.chori.org/bacpac/rat230.htm>). For BAC library



availability, please contact Pieter de Jong (pjejong@mail.cho.org).  
 Clones may be purchased from BACPAC Resources  
 (http://www.choi.org/bacpac/or ering.information.htm). BAC end  
 page: http://www.tigr.org/tldb/bac\_ends/rat/bac\_end\_intro.html  
 Plate: 159 row: A column: 24  
 Seq primer: SP6  
 Class: BAC ends

FEATURES  
 source  
 Location/Qualifiers  
 1..560  
 /organism="Rattus norvegicus"  
 /strain="BN/SSNHsd/MCW"  
 /db\_xref="taxon:10116"  
 /clone="CH230-159A24"  
 /clone\_lib="CHORI-230 Segment 1"  
 /sex="Female"  
 /cell\_type="Brain"  
 /note="Vector: pTARBAC2.1; Site\_1: EcoRI; Site\_2: EcoRI;  
 CHORI-230 Rat (BN/SSNHsd/MCW) BAC library produced by  
 Pieter de Jong"  
 BASE COUNT 101 a 188 c 125 g 146 t  
 ORIGIN

Query Match 4.1%; Score 30; DB 12; Length 560;  
 Best Local Similarity 100.0%; Pred. No. 4.1e+02;  
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 443 CCACCTCTTTTTTTTTTTTTTTTTTTGG 472  
 Db 320 CCACCTCTTTTTTTTTTTTTTTTTTTGG 349

RESULT 14  
 BH316471  
 LOCUS  
 DEFINITION CH230-159E22.TJ CHORI-230 Segment 1 Rattus norvegicus genomic clone  
 CH230-159E22, DNA sequence.  
 ACCESSION BH316471  
 VERSION  
 KEYWORDS  
 SOURCE GSS.  
 ORGANISM Norway rat.  
 Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.  
 1 (bases 1 to 573)  
 Zhao,S., Shetty,J., Shatsman,S., Tsengaye,G., Geer,K., Shvartsbeyn  
 ,A., Gebreyegorgis,E., Overton,L., Russell,D., Chen,D., Riggs,F., de  
 Jong,P. and Fraser,C.M.  
 Rat BAC End Sequences from Library CHORI-230 EcoRI segment  
 Unpublished (1999)  
 Other\_GSSs: CH230-159E22.TV  
 Contact: Shaying Zhao  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: szhao@tigr.org  
 Clones are derived from the rat BAC library CHORI-230  
 (http://www.choi.org/bacpac/rat230.htm). For BAC library  
 availability, please contact Pieter de Jong (pjejong@mail.cho.org).  
 Clones may be purchased from BACPAC Resources  
 (http://www.choi.org/bacpac/or ering.information.htm). BAC end  
 page: http://www.tigr.org/tldb/bac\_ends/rat/bac\_end\_intro.html  
 Plate: 159 row: E column: 22  
 Seq primer: SP6  
 Class: BAC ends.

FEATURES  
 source  
 Location/Qualifiers  
 1..573  
 /organism="Rattus norvegicus"  
 /strain="BN/SSNHsd/MCW"  
 /db\_xref="taxon:10116"

/clone="CH230-159E22"  
 /clone\_lib="CHORI-230 Segment 1"  
 /sex="Female"  
 /cell\_type="Brain"  
 /note="Vector: pTARBAC2.1; Site\_1: EcoRI; Site\_2: EcoRI;  
 CHORI-230 Rat (BN/SSNHsd/MCW) BAC library produced by  
 Pieter de Jong"  
 BASE COUNT 103 a 194 c 129 g 147 t  
 ORIGIN

Query Match 4.1%; Score 30; DB 12; Length 573;  
 Best Local Similarity 100.0%; Pred. No. 4e+02;  
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 443 CCACCTCTTTTTTTTTTTTTTTTTTTGG 472  
 Db 311 CCACCTCTTTTTTTTTTTTTTTTTTTGG 340

RESULT 15  
 AA946504/c  
 LOCUS  
 DEFINITION EST202003 Normalized rat ovary, Bento Soares Rattus sp. cDNA clone  
 ROVA890 3' end, mRNA sequence.  
 ACCESSION AA946504  
 VERSION  
 KEYWORDS  
 SOURCE EST.  
 ORGANISM Rattus sp.  
 Rattus sp.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.  
 1 (bases 1 to 585)  
 Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,  
 Kervavage,A.R. and Adams,M.D.  
 Rat Genome Project: Generation of a Rat EST (RESE) Catalog & Rat  
 Gene Index  
 Unpublished (1998)  
 Contact: Lee, NH  
 The Institute for Genomic Research  
 9712 Medical Center Drive, Rockville, MD 20850, USA  
 Tel: (301)-838-3529  
 Fax: (301)-838-0208  
 Email: nhlee@tigr.org  
 Seq primer: M13-21.  
 Location/Qualifiers  
 1..585  
 /organism="Rattus sp."  
 /db\_xref="ATCC (inhost):2020032"  
 /db\_xref="taxon:10118"  
 /clone="ROVA890"  
 /clone\_lib="Normalized rat ovary, Bento Soares"  
 /note="Organ: ovary; Vector: pT73pac; Site\_1: EcoRI;  
 Site\_2: NotI"  
 BASE COUNT 222 a 98 c 101 g 164 t  
 ORIGIN

Query Match 4.1%; Score 30; DB 9; Length 585;  
 Best Local Similarity 100.0%; Pred. No. 4e+02;  
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 441 AACCACTCTTTTTTTTTTTTTTTTTTTT 470  
 Db 312 AACCACTCTTTTTTTTTTTTTTTTTTTT 283

Search completed: August 2, 2002, 22:41:29  
 Job time: 30144 sec





BASE COUNT 1739 a 1067 c 1319 g 1224 t  
ORIGIN

alignment\_scores:  
Quality: 208.00 Length: 41  
Ratio: 5.073 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-08-973-363-6 x MUSCHDLX ..

Align seg 1/1 to: MUSCHDLX from: 1 to: 5349

1 IleLeuProAspAspProAspLysProGlnAlaLysGlnLeuGlnTh 17  
|||||  
4026 ATTCTCCAGATGATCCTGATATAAAACACAAAGCAAAAGTTACAG 4075  
|||||  
17 rArgAlaAspTyrLeuIleLysLeuSerArgAspLeuAlaLysArg 34  
|||||  
4076 CCGTCGAGACTACCTCATCAAACTACTAGCAGAGATCTTGCAAAAAGAG 4125  
|||||  
34 luAlaGlnArgLeuCysGlyAla 41  
|||||  
4126 AGGCTCAGAGACTTTGTGTGCG 4148  
|||||

seq\_name: gb\_pat:A58683  
seq\_documentation\_block:  
LOCUS A58683 153 bp DNA linear PAT 06-MAR-1998  
DEFINITION Sequence 2 from Patent WO9639505.  
ACCESSION A58683  
VERSION A58683.1 GI:3714246  
KEYWORDS  
SOURCE  
ORGANISM  
unidentified.  
unclassified.  
REFERENCE 1 (bases 1 to 153)  
AUTHORS Griffiths,R. and Tiwari,B.  
TITLE AVIAN GHD GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN BIRDS  
JOURNAL Patent: WO 9639505-A 2 12-DEC-1996;  
COMMENT ISIS INNOVATION (GB)  
FEATURES Other publication AU 5906996 961224.  
Location/Qualifiers  
source  
1. .153  
/organism="unidentified"  
/db\_xref="taxon:32644"

BASE COUNT 61 a 37 c 27 g 28 t  
ORIGIN

alignment\_scores:  
Quality: 193.00 Length: 51  
Ratio: 4.707 Gaps: 1  
Percent Similarity: 80.392 Percent Identity: 80.392

alignment\_block:  
US-08-973-363-6 x A58683 ..

Align seg 1/1 to: A58683 from: 1 to: 153

1 IleLeuProAspAspProAsp..... 7  
|||||  
1 ATTCTCCAGATGATCCTGATATAAAACACAAAGCAAAAGTTACAGAC 50  
|||||  
8 .LysLysProGlnAlaLysGlnLeuGlnThrArgAlaAspTyrLeuIle 24  
|||||  
51 CAAAAAACACAAAGCAAAACAGTTACAGCCCGTCGAGACTACCTCATCA 100  
|||||  
24 ystLeuSerArgAspLeuAlaLysArgGluAlaGlnArgLeuCysGly 40  
|||||  
101 AACTACTTAGCAGAGATCTTGCAAAAGAGAGGCTCAGAGACTTTGTGGT 150  
|||||

41 Ala 41  
|||  
151 GCG 153

seq\_name: gb\_pat:A58696

seq\_documentation\_block:

LOCUS A58696 1316 bp DNA linear PAT 06-MAR-1998  
DEFINITION Sequence 15 from Patent WO9639505.  
ACCESSION A58696  
VERSION A58696.1 GI:3714253  
KEYWORDS  
SOURCE  
ORGANISM  
unidentified.  
unclassified.  
REFERENCE 1 (bases 1 to 1316)  
AUTHORS Griffiths,R. and Tiwari,B.  
TITLE AVIAN GHD GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN BIRDS  
JOURNAL Patent: WO 9639505-A 15 12-DEC-1996;  
COMMENT ISIS INNOVATION (GB)  
FEATURES Other publication AU 5906996 961224.  
Location/Qualifiers  
source  
1. .1316  
/organism="unidentified"  
/db\_xref="taxon:32644"

BASE COUNT 493 a 205 c 308 g 304 t 6 others  
ORIGIN

alignment\_scores:  
Quality: 187.00 Length: 41  
Ratio: 4.561 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 87.805

alignment\_block:  
US-08-973-363-6 x A58696 ..

Align seg 1/1 to: A58696 from: 1 to: 1316

1 IleLeuProAspAspProAspLysLysProGlnAlaLysGlnLeuGlnTh 17  
|||||  
1138 ATTTTACCTGATGATCCAGATAGAAACCCAGGCTAAGCAGTTACAGAC 1187  
|||||  
17 rArgAlaAspTyrLeuIleLysLeuSerArgAspLeuAlaLysArg 34  
|||||  
1188 TCGTCGAGATTACCTATTAAATTACTGAATAAGACCTTGCAAGAAAGG 1237  
|||||  
34 luAlaGlnArgLeuCysGlyAla 41  
|||||  
1238 AAGCACAGAGACTTGCTGTGCA 1260  
|||||

seq\_name: gb\_pat:A58691

seq\_documentation\_block:

LOCUS A58691 6608 bp DNA linear PAT 06-MAR-1998  
DEFINITION Sequence 10 from Patent WO9639505.  
ACCESSION A58691  
VERSION A58691.1 GI:3714250  
KEYWORDS  
SOURCE  
ORGANISM  
unidentified.  
unclassified.  
REFERENCE 1 (bases 1 to 6608)  
AUTHORS Griffiths,R. and Tiwari,B.  
TITLE AVIAN GHD GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN BIRDS  
JOURNAL Patent: WO 9639505-A 10 12-DEC-1996;  
COMMENT ISIS INNOVATION (GB)  
FEATURES Other publication AU 5906996 961224.  
Location/Qualifiers  
source  
1. .6608





seq\_documentation\_block:  
 LOCUS AC012624 134365 bp DNA linear PRI 21-JUL-2001  
 DEFINITION Homo sapiens chromosome 5 clone CTD-2082117, complete sequence.  
 ACCESSION AC012624  
 VERSION AC012624.6 GI:14993679  
 KEYWORDS HTG  
 SOURCE human  
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 134365)  
 DOE Joint Genome Institute and Stanford Human Genome Center.  
 Direct Submission  
 Unpublished  
 2 (bases 1 to 134365)  
 DOE Joint Genome Institute.  
 Direct Submission  
 3 (bases 1 to 134365)  
 Submitted (31-OCT-1999) Production Sequencing Facility, DOE Joint  
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
 4 (bases 1 to 134365)  
 Submitted (01-JUN-2001) DOE Joint Genome Institute, 2800 Mitchell  
 Drive, Walnut Creek, CA 94598, USA  
 Direct Submission  
 5 (bases 1 to 134365)  
 Submitted (21-JUL-2001) DOE Joint Genome Institute, 2800 Mitchell  
 Drive, Walnut Creek, CA 94598, USA  
 On Jul 21, 2001 this sequence version replaced gi:14277267.  
 Location/Qualifiers  
 1. 134365

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

\* 75509 92516: contig of 17008 bp in length  
\* 92517 92616: gap of 100 bp  
\* 92617 106409: contig of 13793 bp in length  
\* 106410 106509: gap of 100 bp  
\* 106510 143079: contig of 36570 bp in length.

## FEATURES

source  
1. .143079  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="RP11-S8M12"  
/clone\_lib="RPC1-11 Human Male BAC"  
1. .38820  
/note="assembly\_fragment"

## misc\_feature

clone\_end:SP6  
vector\_side:left  
38921..40411  
/note="assembly\_fragment"

## misc\_feature

40512..43279  
/note="assembly\_fragment"

## misc\_feature

43380..46905  
/note="assembly\_fragment"

## misc\_feature

47006..51830  
/note="assembly\_fragment"

## misc\_feature

51931..62619  
/note="assembly\_fragment"

## misc\_feature

62720..75408  
/note="assembly\_fragment"

## misc\_feature

75509..92516  
/note="assembly\_fragment"

## misc\_feature

92617..106409  
/note="assembly\_fragment"

## misc\_feature

106510..143079  
/note="assembly\_fragment"

## misc\_feature

clone\_end:T7  
vector\_side:right

BASE COUNT 43971 a 26246 c 26678 g 45278 t 906 others  
ORIGIN

## alignment\_scores:

Quality: 174.00 Length: 41  
Ratio: 4.579 Gaps: 1  
Percent Similarity: 92.683 Percent Identity: 90.244

## alignment\_block:

US-08-973-363-6 x AC021449 ..

Align seg 1/1 to: AC021449 from: 1 to: 143079

1 IleLeuProAspAspProAspLysProGlnAlaLysGlnLeuGln 17  
|||||  
116147 ATCTTCCAGATGATCCCATATAAAACCAACAGCAACAGTTCGAGAC 116196

17 rArgAlaAspTyrLeuIleLysLeuLeuSerArgAspLeuAlaLysArg 34  
|||||  
116197 CCGTCGACACTACCTCATCAAAATTACTTAGTAGAGATCTTGCAAAAAG 116246

34 luAlaGlnArgLeuCysGlyAla 41  
|||||

116247 AGCT.....CTTCTGTGTCG 116263

seq\_name: gb\_htg:AC008531

seq\_documentation\_block:

LOCUS AC008531 145659 bp DNA linear HTG 14-FEB-2001  
DEFINITION Homo sapiens chromosome 5 clone CTC-480B11, WORKING DRAFT SEQUENCE,  
7 ordered pieces.

ACCESSION AC008531

VERSION AC008531.3 GI:12830078

KEYWORDS HTG; HTGS\_PHASE2; HTGS\_DRAFT.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 145659)  
DOE Joint Genome Institute.  
Sequencing of Human Chromosome 5  
Unpublished  
2 (bases 1 to 145659)  
DOE Joint Genome Institute.  
Direct Submision  
Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
On Feb 14, 2001 this sequence version replaced gi:7528342.  
-----Genome Center  
Center: Joint Genome Institute  
Center Code: JGI  
Web site: http://www.jgi.doe.gov  
-----

## Project Information

Center Project Name: 369535  
Center clone name: CIT-HSPC\_480B11  
-----

## Summary Statistics

Consensus quality: 139128 bases at least Q40  
Consensus quality: 142556 bases at least Q30  
Consensus quality: 143744 bases at least Q20  
Estimated insert size: 148000; pulse field gel estimation  
Estimated insert size: 145059; sum-of-contigs estimation  
Quality coverage: 6.27 in Q20 bases; pulse field gel estimation  
Quality coverage: 6.4 in Q20 bases; sum-of-contigs estimation.  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 7 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* provided by the submittor.

\* This sequence will be replaced

\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.

\* 1 56174: contig of 56174 bp in length  
\* 56175 56274: gap of unknown length  
\* 56275 100874: contig of 44600 bp in length  
\* 100875 100974: gap of unknown length  
\* 100975 113127: contig of 12153 bp in length  
\* 113128 113227: gap of unknown length  
\* 113228 118190: contig of 4963 bp in length  
\* 118191 118290: gap of unknown length  
\* 118291 118694: contig of 1404 bp in length  
\* 118695 119794: gap of unknown length  
\* 119795 123297: contig of 3503 bp in length  
\* 123298 123398: gap of unknown length  
\* 123398 145659: contig of 22262 bp in length.

FEATURES  
source

1. .145659  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="5"  
/clone="CTC-480B11"  
/clone\_lib="CalTech human BAC library C"  
BASE COUNT 42561 a 26309 c 27580 g 48609 t 600 others  
ORIGIN

## alignment\_scores:

Quality: 174.00 Length: 41  
Ratio: 4.579 Gaps: 1  
Percent Similarity: 92.683 Percent Identity: 90.244

## alignment\_block:

US-08-973-363-6 x AC008531 ..

Align seg 1/1 to: AC008531 from: 1 to: 145659

1 IleLeuProAspAspProAspLysProGlnAlaLysGlnLeuGln 17  
|||||



46770 ATTCTCCAGATGATCCGATATAAAAAACACAGCAAAACAGTTGCAGAC 46819

17 ratgalaasptvrlleuileuylsleuSerArcaAspLeuAlaLysArgG 34  
|||||

46820 CCGTGCAGACTACCTCACTCAAAATTAAGTAGAGATCTTGCACAAAAG 46869

34 luAlaGlnArgLeuCysGlyAla 41  
|||||

46870 AGGCT.....CTTCTGGTGGC 46886

seq\_name: gb\_htg:AC091946

seq\_documentation\_block:

LOCUS AC091946 193446 bp DNA linear HTG 09-JUN-2001  
DEFINITION Homo sapiens chromosome 5 clone RP11-36012, \*\*\* SEQUENCING IN  
PROGRESS \*\*\*, 33 unordered pieces.

ACCESSION AC091946

VERSION AC091946.1 GI:14333882

KEYWORDS HTG; HTGS\_PHASE1.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1. (bases 1 to 193446)

DOE Joint Genome Institute.

Sequencing of Human Chromosome 5

Unpublished

2. (bases 1 to 193446)

DOE Joint Genome Institute.

Direct Submission

Submitted (09-JUN-2001) Production Sequencing Facility, DOE Joint

Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

-----Genome Center

Center: Joint Genome Institute

Center Code: JGI

Web site: http://www.jgi.doe.gov

-----

Project Information

Center Project Name: 544799

Center clone name: RPCI-11\_36012

-----

Summary Statistics

Consensus quality: 157767 bases at least Q40

Consensus quality: 180259 bases at least Q30

Consensus quality: 184175 bases at least Q20

Estimated insert size: 204530; agarose-fp estimation

Estimated insert size: 190246; sum-of-contigs estimation

Quality coverage: 5.38 in Q20 bases; agarose-fp estimation

Quality coverage: 5.79 in Q20 bases; sum-of-contigs estimation.

\* NOTE: this is a 'working draft' sequence. It currently

\* consists of 33 contigs. The true order of the pieces

\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence

\* as soon as it is available and the accession number will

\* be preserved.

1 1358: contig of 1358 bp in length

\* 1359 1458: gap of unknown length

\* 1459 2667: contig of 1209 bp in length

\* 2668 2767: gap of unknown length

\* 2768 3823: contig of 1056 bp in length

\* 3824 3923: gap of unknown length

\* 3924 5920: contig of 1997 bp in length

\* 5921 6020: gap of unknown length

\* 6021 7494: contig of 1474 bp in length

\* 7495 7594: gap of unknown length

\* 7595 9096: contig of 1502 bp in length

\* 9097 9197: gap of unknown length

\* 9197 10260: contig of 1064 bp in length

\* 10261 10360: gap of unknown length

\* 10361 12460: contig of 2100 bp in length

\* 12461 12560: gap of unknown length

\* 12561 14611: contig of 2051 bp in length

\* 14612 14711: gap of unknown length

\* 14712 16381: contig of 1670 bp in length

\* 16382 16481: gap of unknown length

\* 16482 17958: contig of 1487 bp in length

\* 17959 18058: gap of unknown length

\* 18059 20434: contig of 2366 bp in length

\* 20435 20535: gap of unknown length

\* 20536 23516: contig of 2981 bp in length

\* 23517 27563: gap of unknown length

\* 27564 27663: contig of 3948 bp in length

\* 27664 30987: gap of unknown length

\* 30988 31087: contig of 3324 bp in length

\* 31088 36065: contig of 4978 bp in length

\* 36066 36165: gap of unknown length

\* 36166 40978: contig of 4813 bp in length

\* 40979 41078: gap of unknown length

\* 41079 45663: contig of 4585 bp in length

\* 45664 45763: gap of unknown length

\* 45764 51745: contig of 5982 bp in length

\* 51746 51845: gap of unknown length

\* 51846 57359: contig of 5514 bp in length

\* 57360 57459: gap of unknown length

\* 57460 67881: contig of 10422 bp in length

\* 67882 67981: gap of unknown length

\* 67982 74132: contig of 6151 bp in length

\* 74133 74232: gap of unknown length

\* 74233 79795: contig of 5563 bp in length

\* 79796 79895: gap of unknown length

\* 79896 87511: contig of 7616 bp in length

\* 87512 87611: gap of unknown length

\* 87612 92891: contig of 5180 bp in length

\* 92892 92891: gap of unknown length

\* 92892 102794: contig of 9903 bp in length

\* 102795 102894: gap of unknown length

\* 102895 110866: contig of 7972 bp in length

\* 110867 110966: gap of unknown length

\* 110967 117571: contig of 6605 bp in length

\* 117572 117571: gap of unknown length

\* 117572 123738: contig of 6067 bp in length

\* 123739 123838: gap of unknown length

\* 123839 130583: contig of 6745 bp in length

\* 130584 130683: gap of unknown length

\* 130684 141544: contig of 10861 bp in length

\* 141545 141644: gap of unknown length

\* 141645 169109: contig of 27465 bp in length

\* 169110 169209: gap of unknown length

\* 169210 193446: contig of 24237 bp in length.

#### FEATURES

source

1. 193446

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="5"

/clone="RP11-36012"

/clone\_lib="RPCI human BAC library 11"

BASE COUNT 57571 a 35252 c 37061 g 60360 t 3202 others

ORIGIN

alignment\_scores:

Quality: 174.00

Ratio: 4.579

Percent Similarity: 92.683

Length: 41

Gaps: 1

Percent Identity: 90.244

alignment\_block:

US-08-973-363-6 x AC091946

Align seg 1/1 to: AC091946 from: 1 to: 193446

1 lleuProaspPProaspLysProGlnAlaLysGlnInth 17

|||||

|||||

|||||

|||||

|||||

|||||

|||||

87658 ATTCTCCAGATGATCCGATATAAAAAACACAGCAAAACAGTTGCAGAC 87707

17 rArgAlaAspTyrLeuIleLysLeuSerArgAspLeuAlaLysArgG 34  
|||||  
87708 CCGTGCAGACTACCTCATCAAAATTACTAGTAGAGATCTTGCAAAAAAG 87757

34 luAlaGlnArgLeuCysGlyAla 41  
|||||  
87758 AAGCTT.....CTTCTGGTGCG 87774

seq\_name: gb\_pr:AC026778

seq\_documentation\_block:  
LOCUS AC026778 195433 bp DNA linear PRI 01-JUN-2001  
DEFINITION Homo sapiens chromosome 5 clone CTC-428111, complete sequence.

ACCESSION AC026778  
VERSION AC026778.4 GI:14277282

KEYWORDS HTG.

SOURCE human.

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 195433)

AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.

TITLE Direct Submission

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 195433)

AUTHORS DOE Joint Genome Institute.

TITLE Direct Submission

JOURNAL Submitted (24-MAR-2000) Production Sequencing Facility, DOE Joint

Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

REFERENCE 3 (bases 1 to 195433)

AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.

TITLE Direct Submission

JOURNAL Submitted (01-JUN-2001) DOE Joint Genome Institute, 2800 Mitchell

Drive, Walnut Creek, CA 94598, USA

COMMENT On Jun 1, 2001 this sequence version replaced gi:13677045.

Draft Sequence Produced by DOE Joint Genome Institute

www.jgi.doe.gov

Finishing Completed at Stanford Human Genome Center

www.shgc.stanford.edu

Quality: Phrap Quality >=40 99.9% of Sequence;

Estimated Total Number of Errors is 0.2.

STS Content:

WI-13675 G23101

SHGC-58345 G38487

SHGC-103595 G57841.

FEATURES Location/Qualifiers

source 1..195433

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="5"

/clone="CTC-428111"

BASE COUNT 62762 a 37302 c 37040 g 58329 t

ORIGIN

alignment\_scores:

Quality: 174.00 Length: 41

Ratio: 4.579 Gaps: 1

Percent Similarity: 92.683 Percent Identity: 90.244

alignment\_block:

US-08-973-363-6 x AC026778/rev ..

Align seg 1/1 to reverse of: AC026778 from: 1 to: 195433

1 IleLeuProAspProAspLysProGlnAlaLysGlnLeuGlnTh 17

|||||

29732 ATTCTCCAGATGATCCCGATAAAAACCAACAGCAAGTTCGAGAC 29683

|||||

17 rArgAlaAspTyrLeuIleLysLeuSerArgAspLeuAlaLysArgG 34

|||||

29682 CCGTGCAGACTACCTCATCAAAATTACTAGTAGAGATCTTGCAAAAAAG 29633

34 luAlaGlnArgLeuCysGlyAla 41  
|||||  
29632 AAGCTT.....CTTCTGGTGCG 29616

seq\_name: gb\_pr:AC022121

seq\_documentation\_block:

LOCUS AC022121

DEFINITION Homo sapiens chromosome 5 clone CTD-2007H13, complete sequence.

ACCESSION AC022121

VERSION AC022121.6 GI:15375145

KEYWORDS HTG.

SOURCE human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 219258)

AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.

TITLE Direct Submission

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 219258)

AUTHORS DOE Joint Genome Institute.

TITLE Direct Submission

JOURNAL Submitted (26-JAN-2000) Production Sequencing Facility, DOE Joint

Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

REFERENCE 3 (bases 1 to 219258)

AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.

TITLE Direct Submission

JOURNAL Submitted (11-AUG-2001) DOE Joint Genome Institute, 2800 Mitchell

Drive, Walnut Creek, CA 94598, USA

COMMENT On Aug 30, 2001 this sequence version replaced gi:15148108.

Draft Sequence Produced by DOE Joint Genome Institute

www.jgi.doe.gov

Finishing Completed at Stanford Human Genome Center

www.shgc.stanford.edu

Quality: Phrap Quality >=40 99.8% of Sequence;

Estimated Total Number of Errors is 0.4.

STS Content:

WI-5811 G04974

SHGC-58345 G38487

SHGC-103595 G57841.

FEATURES Location/Qualifiers

source 1..219258

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="5"

/clone="CTD-2007H13"

BASE COUNT 71954 a 42062 c 40933 g 64309 t

ORIGIN

alignment\_scores:

Quality: 174.00 Length: 41

Ratio: 4.579 Gaps: 1

Percent Similarity: 92.683 Percent Identity: 90.244

alignment\_block:

US-08-973-363-6 x AC022121/rev ..

Align seg 1/1 to reverse of: AC022121 from: 1 to: 219258

1 IleLeuProAspProAspLysProGlnAlaLysGlnLeuGlnTh 17

|||||

89450 ATTCTCCAGATGATCCCGATAAAAACCAACAGCAAGTTCGAGAC 89401

|||||

17 rArgAlaAspTyrLeuIleLysLeuSerArgAspLeuAlaLysArgG 34

```
|||||
89400 CCGTGCAGACTACTCATCAATTAATTAGTAGAGATCTTGCAGAAAAAG 89351
34 luLaGlnArgLeuGlyAla 41
|||||
89350 AGCT.....CTTCTGGTGGCG 89334

seq_name: gb_htg.AC092382

seq_documentation_block:
LOCUS AC092382 . 276181 bp DNA linear HTG 03-JUL-2001
DEFINITION Homo sapiens chromosome 5 clone RF11-75H1, WORKING DRAFT SEQUENCE,
47 unprocessed pieces.
ACCESSION AC092382
VERSION AC092382.1 GI:14589571
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 276181)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 5
JOURNAL Unpublished
REFERENCE
2 (bases 1 to 276181)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (03-JUL-2001) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell drive, Walnut Creek, CA 94598, USA
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 435334
Center clone name: RPCI-11_75H1
-----
Summary Statistics
Consensus quality: 229677 bases at least Q40
Consensus quality: 256163 bases at least Q30
Consensus quality: 260799 bases at least Q20
Estimated insert size: 174820; agarose-fp estimation
Estimated insert size: 271581; sum-of-contigs estimation
Quality coverage: 8.62 in Q20 bases; agarose-fp estimation
Quality coverage: 5.95 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 47 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1061: contig of 1061 bp in length
* 1062 1161: gap of unknown length
* 1162 2827: contig of 1666 bp in length
* 2828 2927: gap of unknown length
* 2928 4227: contig of 1300 bp in length
* 4228 4327: gap of unknown length
* 4328 5618: contig of 1291 bp in length
* 5619 5718: gap of unknown length
* 5719 6983: contig of 1265 bp in length
* 6984 7083: gap of unknown length
* 7084 8422: contig of 1339 bp in length
* 8423 8522: gap of unknown length
* 8523 9771: contig of 1249 bp in length
* 9772 9872: gap of unknown length
* 9873 11199: contig of 1327 bp in length
* 11199 12154: gap of unknown length
* 12155 12515: contig of 1116 bp in length
* 12516 13806: gap of unknown length
* 13807 13906: gap of unknown length
* 13907 15360: contig of 1454 bp in length
* 15361 15460: gap of unknown length
* 15461 16687: contig of 1207 bp in length
* 16688 16787: gap of unknown length
* 16788 18390: contig of 1623 bp in length
* 18391 18491: gap of unknown length
* 18492 19541: contig of 1051 bp in length
* 19542 19641: gap of unknown length
* 19642 21156: contig of 1515 bp in length
* 21157 21257: gap of unknown length
* 21258 22548: contig of 1192 bp in length
* 22549 24599: contig of 1951 bp in length
* 24600 26384: contig of 1785 bp in length
* 26385 26484: gap of unknown length
* 26485 27883: contig of 1399 bp in length
* 27884 27983: gap of unknown length
* 27984 30386: contig of 2403 bp in length
* 30387 30486: gap of unknown length
* 30487 32580: contig of 1994 bp in length
* 32581 35441: contig of 2861 bp in length
* 35442 37800: contig of 2259 bp in length
* 37801 37900: gap of unknown length
* 37901 40587: contig of 2687 bp in length
* 40588 40687: gap of unknown length
* 40688 44159: contig of 3472 bp in length
* 44160 44259: gap of unknown length
* 44260 46636: contig of 2377 bp in length
* 46637 46737: gap of unknown length
* 46738 50082: contig of 3346 bp in length
* 50083 50182: gap of unknown length
* 50183 53988: contig of 3806 bp in length
* 53989 54088: gap of unknown length
* 54089 56592: contig of 2504 bp in length
* 56593 61352: gap of unknown length
* 61353 61452: contig of 4660 bp in length
* 61453 64254: gap of unknown length
* 64255 64354: contig of 2802 bp in length
* 64355 68255: gap of unknown length
* 68256 68925: contig of 4471 bp in length
* 68926 74393: gap of unknown length
* 74394 80561: contig of 5468 bp in length
* 80562 80661: gap of unknown length
* 80662 87626: contig of 6068 bp in length
* 87627 87726: gap of unknown length
* 87727 93600: contig of 6965 bp in length
* 93601 93701: gap of unknown length
* 93702 102024: contig of 5874 bp in length
* 102025 102124: gap of unknown length
* 102125 109905: contig of 8324 bp in length
* 109906 110005: contig of 7781 bp in length
* 110006 116737: gap of unknown length
* 116738 116837: contig of 6732 bp in length
* 116838 126797: gap of unknown length
* 126798 126897: contig of 9960 bp in length
* 126898 135114: gap of unknown length
* 135115 135214: contig of 8217 bp in length
* 135215 146713: gap of unknown length
* 146714 146813: contig of 11499 bp in length
* 146814 157894: gap of unknown length
* 157895 171752: contig of 11081 bp in length
* 171753 171852: gap of unknown length
* 171853 201687: contig of 13756 bp in length
* 201688 201788: gap of unknown length
* 201789 231268: contig of 29835 bp in length
* 231269 231368: gap of unknown length
```

\* 231369 276181: contig of 44813 bp in length.  
FEATURES  
    Location/Qualifiers  
        1..276181  
            /organism="Homo sapiens"  
            /db\_xref="taxon:9606"  
            /chromosome="5"  
            /clone="RP11-75H1"  
            /clone\_lib="RPCI human BAC library 11"  
BASE COUNT 86859 a 51769 c 50026 g 82854 t 4673 others  
ORIGIN

alignment\_scores:  
    Quality: 174.00           Length: 41  
    Ratio: 4.579           Gaps: 1  
Percent Similarity: 92.683   Percent Identity: 90.244

alignment\_block:

US-08-973-363-6 x AC092382/rev ..

Align seg 1/1 to reverse of: AC092382 from: 1 to: 276181

1 IleLeuProAspAspProAspLysLysProGlnAlaLysGlnLeuGlnTh 17  
|||||  
213694 ATCTTCCAGATGATCCGATATAAAACACAAAGCAAAACAGTTGCAGAC 213645

17 rArgAlaAspTyrLeuIleLysLeuSerArgAspLeuAlaLysArgG 34  
|||||  
213644 CCGTGCACTACTCATCAATATTACTTAGTAGAGATCTTGCAAAAAAAG 213595

34 luAlaGlnArgLeuCysGlyAla 41

|||||  
213594 AAGCT.....CTTCTGGTGCG 213578

seq\_name: gb\_pat:A58684

seq\_documentation\_block:  
LOCUS A58684 153 bp DNA linear PAT 06-MAR-1998  
DEFINITION Sequence 3 from Patent WO9639505.  
ACCESSION A58684  
VERSION A58684.1 GI:3714247  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 153)  
AUTHORS Griffiths R. and Tiwari, B.  
TITLE AVIAN GHD GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN BIRDS  
JOURNAL Patent: WO 9639505-A 3 12-DEC-1996;  
COMMENT ISIS INNOVATION (GB)  
FEATURES Other publication AU 5906996 961224.  
          Location/Qualifiers  
          1..153  
          /organism="unidentified"  
          /db\_xref="taxon:32644"  
BASE COUNT 58 a 40 c 31 g 24 t  
ORIGIN

alignment\_scores:  
    Quality: 172.00           Length: 51  
    Ratio: 4.195           Gaps: 1  
Percent Similarity: 80.392   Percent Identity: 70.588

alignment\_block:

US-08-973-363-6 x A58684 ..

Align seg 1/1 to: A58684 from: 1 to: 153

1 IleLeuProAspAspProAsp.....7  
|||||  
1 ATTTTACCTGATGATCCAGACAGAAACCCAGGCAAGCAGCTACAGAC 50

8 .LysLysProGlnAlaLysGlnLeuGlnThrArgAlaAspTyrLeuIleL 24  
|||||  
51 CAAGAAACCCAGGCAAGCAGCTACAGACCCGTCAGACTACTCATTA 100  
|||||  
24 ysLeuLeuSerArgAspLeuAlaLysArgGluAlaGlnArgLeuCysGly 40  
|||||  
101 AATTACTGAATAAAGACCTTGCAAGAAAGGAGCACAAGGCTTGCTGGT 150  
|||||  
41 Ala 41  
151 GCA 153

Mon Aug 5 11:51:56 2002

us-08-973-363-6.p2n.rge

Page 11

**THIS PAGE BLANK (USPTO)**

OM of: US-08-973-363-6 to: N\_Geneseq\_032802.\* out\_format : pfs  
Date: Aug 3, 2002 4:34 AM  
About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 Compugen Ltd.

## Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh  
-O=cpn2.1.uspto.scol\US08973363\runat\_01082002\_080050\_18606\app\_query.fasta\_1.638  
-DB=N\_Geneseq\_032802 -QFMT=fastap -SUFFIX=p2n.rng -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOQCL=0.000 -LOQEXT=0.000  
-CGAPOP=4.500 -CGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=biosum62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR\_SCORE=pct  
-THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs  
-NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US08973363 -CGN1\_1.186 -NCPU=6 -ICPU=3 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -NO\_XLPXY -WAIT -THREADS=1

## Search information block:

Query: US-08-973-363-6  
Query length: 41  
Database: N\_Geneseq\_032802.\*  
Database sequences: 1736436  
Database length: 858457221  
Search time (sec): 523.620000

## score\_list:

Sequence	Strd Orig	ZScore	Escore	Len	Documentation
/SIDSI/gcgdata/hold-geneseq/geneseq-nbml/NA1997.DAT:AA742756 +		193.00	538.46	1.2e-21	
/SIDSI/gcgdata/hold-geneseq/geneseq-nbml/NA1997.DAT:AA742754 +		187.00	500.54	1.6e-19	
/SIDSI/gcgdata/hold-geneseq/geneseq-nbml/NA1997.DAT:AA742751 +		187.00	484.74	1.2e-18	
/SIDSI/gcgdata/hold-geneseq/geneseq-nbml/NA1997.DAT:AA742758 +		172.00	479.52	2.3e-18	
/SIDSI/gcgdata/hold-geneseq/geneseq-nbml/NA1997.DAT:AA742759 +		172.00	479.52	2.3e-18	
/SIDSI/gcgdata/hold-geneseq/geneseq-nbml/NA1998.DAT:AA742759 +		167.00	465.49	1.4e-17	
/SIDSI/gcgdata/hold-geneseq/geneseq-nbml/NA2001A.DAT:AAK88888 +		129.00	337.81	1.8e-10	
/SIDSI/gcgdata/hold-geneseq/geneseq-nbml/NA2001A.DAT:AA757603 +		118.00	318.06	2.3e-0	
/SIDSI/gcgdata/hold-geneseq/geneseq-nbml/NA2001A.DAT:AB106443 +		84.50	197.65	0.011e	
/SIDSI/gcgdata/hold-geneseq/geneseq-nbml/NA2001B.DAT:AB106442 +		84.50	193.09	0.020e	
/SIDSI/gcgdata/hold-geneseq/geneseq-nbml/NA2002.DAT:AA756071 +		60.00	140.50	17.73	
/SIDSI/gcgdata/hold-geneseq/geneseq-nbml/NA2001A.DAT:AA756036 +		58.00	124.89	131.13	
/SIDSI/gcgdata/hold-geneseq/geneseq-nbml/NA2001A.DAT:AA756032 +		58.00	106.85	1.4e-0	
/SIDSI/gcgdata/hold-geneseq/geneseq-nbml/NA2001B.DAT:AB12760 +		57.00	128.84	78.99	
/SIDSI/gcgdata/hold-geneseq/geneseq-nbml/NA1999.DAT:AA75679 +		57.00	127.94	88.71	
/SIDSI/gcgdata/hold-geneseq/geneseq-nbml/NA2001B.DAT:AB126281 +		56.00	132.75	47.89	
/SIDSI/gcgdata/hold-geneseq/geneseq-nbml/NA2001B.DAT:AB126278 +		56.00	124.17	143.82	
/SIDSI/gcgdata/hold-geneseq/geneseq-nbml/NA2001B.DAT:AB126280 +		55.00	123.56	155.50	
/SIDSI/gcgdata/hold-geneseq/geneseq-nbml/NA2000.DAT:AA756315 +		55.00	123.45	157.74	
/SIDSI/gcgdata/hold-geneseq/geneseq-nbml/NA2001A.DAT:AA756830 +		54.50	139.01	21.44	
/SIDSI/gcgdata/hold-geneseq/geneseq-nbml/NA2001A.DAT:AA756830 +		54.50	139.01	21.44	
/SIDSI/gcgdata/hold-geneseq/geneseq-nbml/NA2001A.DAT:AA756830 +		54.50	139.01	21.44	
/SIDSI/gcgdata/hold-geneseq/geneseq-nbml/NA2001B.DAT:AA756831 +		54.50	139.01	21.44	
/SIDSI/gcgdata/hold-geneseq/geneseq-nbml/NA2002.DAT:AA756831 +		54.50	120.87	219.67	
/SIDSI/gcgdata/hold-geneseq/geneseq-nbml/NA1999.DAT:AA756831 +		54.50	119.11	275.37	
/SIDSI/gcgdata/hold-geneseq/geneseq-nbml/NA2000.DAT:AA756831 +		54.00	126.65	104.52	
/SIDSI/gcgdata/hold-geneseq/geneseq-nbml/NA1999.DAT:AA756831 +		53.50	121.85	193.79	
/SIDSI/gcgdata/hold-geneseq/geneseq-nbml/NA1999.DAT:AA756831 +		53.50	117.72	328.81	
/SIDSI/gcgdata/hold-geneseq/geneseq-nbml/NA2001A.DAT:AA756831 +		53.50	117.64	332.36	
/SIDSI/gcgdata/hold-geneseq/geneseq-nbml/NA2001A.DAT:AA756831 +		53.50	95.90	5.4e-0	
/SIDSI/gcgdata/hold-geneseq/geneseq-nbml/NA1994.DAT:AA756831 +		53.00	128.54	82.12	
/SIDSI/gcgdata/hold-geneseq/geneseq-nbml/NA2001A.DAT:AA756831 +		52.50	118.58	294.51	
/SIDSI/gcgdata/hold-geneseq/geneseq-nbml/NA2001B.DAT:AB105614 +		52.50	107.79	1.2e-0	
/SIDSI/gcgdata/hold-geneseq/geneseq-nbml/NA2001B.DAT:AB121196 +		52.50	107.79	1.2e-0	

/SIDSI/gcgdata/hold-geneseq/geneseq-nbml/NA2001A.DAT:AA756831 + 52.50 90.06 1.1  
/SIDSI/gcgdata/hold-geneseq/geneseq-nbml/NA2001A.DAT:AA756831 + 52.50 68.41 1.8  
/SIDSI/gcgdata/hold-geneseq/geneseq-nbml/NA2002.DAT:AB134316 + 52.00 121.71 197.  
/SIDSI/gcgdata/hold-geneseq/geneseq-nbml/NA2001B.DAT:AA756831 + 52.00 120.60 227  
/SIDSI/gcgdata/hold-geneseq/geneseq-nbml/NA2001B.DAT:AA756831 + 52.00 120.44 232

seq\_name: /SIDSI/gcgdata/hold-geneseq/geneseq-nbml/NA1997.DAT:AA742756

## seq\_documentation\_block:

ID\_AAT42756 standard; DNA: 153 BP.  
AC AAT42756;  
DT 12-MAR-1997 (first entry)  
DE Mouse CHD-1 gene (bases 3855-977).  
KW Bird; sex determination; chromodomain-Helicase-DNA binding 1;  
KW CHD-1; CHD-W; W chromosome; ss.  
OS Mus sp.  
FH Key Location/Qualifiers  
FT misc\_difference 52..81 /\*tag= a  
FT FT /\*note= "bases 52-81 are a repeat of bases 22-51  
FT FT and are ignored in the translated amino  
FT FT acid sequence given in Fig 3"  
PW WO9639505-A1.  
XX 12-DEC-1996.  
XX 05-JUN-1996; 96WO-G801341.  
XX 06-JUN-1995; 95GB-0011439.  
(ISIS-) ISIS INNOVATION LTD.  
XX Griffiths R, Tiwari B;  
XX WPI: 1997-043127/04.  
XX P-PSDB; AAW08146.  
XX Avian chromodomain-helicase-DNA binding genes determine sex in  
birds - used for sex determ. and to control sex of progeny  
XX Claim 8; Fig 3; 76pp; English.  
XX Bases 3855-3977 (AAT42756) of the mouse CHD-1 gene show homology  
to portions of the chicken CHD-1A (A - Avian) gene (see also  
CC and AAT42757), chicken CHD-W (W refers to the W chromosome) gene  
CC (see also AAT42758) and the great tit CHD-W gene (see also AAT42759).  
CC Translated amino acid sequences of this region are provided in  
CC AAW08146-49. The CHD-1A (see also AAT42751) and CHD-W (see also  
CC AAT42754-55) genes determine sex in birds and can be used to identify  
CC the sex of an embryo, foetus etc. and to manipulate the sex of  
progeny.  
XX Sequence 153 BP; 61 A; 37 C; 27 G; 28 T; 0 other;

alignment\_scores:  
Quality: 193.00 Length: 51  
Ratio: 4.707 Gaps: 1  
Percent Similarity: 80.392 Percent Identity: 80.392

## alignment\_block:

US-08-973-363-6 x AAT42756 ..  
Align seg 1/1 to: AAT42756 from: 1 to: 153  
1 lleLeuProAspProAsp..... 7

1 ATTCTTCAGATGATCCTGATAAAAAACCAACGCAAAACAGTTACAGAC 50  
8 .LysLysProGlnAlaLysGlnLeuGlnThrArgAlaAspTyrLeuIleU 24  
51 CAAAAACCCACCAAGCAAAACAGTTACAGACCCGTGCAGACTACCTCATCA 100  
24 ysLeuLeuSerArgAspLeuAlaLysArgGlnAlaGlnArgLeuCysGly 40  
101 AACTACTTACAGAGATCTTCGAAAAAGAGAGGCTCAGAGACITTTGTGT 150  
41 Ala 41  
151 GCG 153

seq\_name: /SIDS1/qcadata/hold-qeneseq/qeneseq-emb1/NA1997.DAT:AAT42754

seq documentation block:

seq\_documentation\_block.  
ID AAT42754 standard: cDNA: 1316 BP.

XX	
XX	
AC	AAT42754;
XX	
DT	12-MAR-1997 (first entry)
XX	
DE	Chicken CHD-W gene (partial sequence).
XX	
KW	Bird; sex determination; chromodomain-Helicase-DNA binding 1 Avian;
KW	CHD-W; chromodomain-Helicase-DNA binding on the W chromosome; ss.

```
alignment_scores:
  Quality: 187.00
  Ratio: 4.561
  Percent Similarity: 100.000
  Percent Identity: 87.805
  Gaps: 0
  Length: 41
```

alignment\_block:  
US-08-973-363-6 x AAT42754

Align seq 1/1 to: AAT42754 from: 1 to: 1316

1 lleLeuProAspAspProAspLysLysProGlnAlaLysGlnLeuGlnTh 17  
1138 ATTTTACCTGATGATCCAGATAAGAAACCCAGAGCTTAAGCAGTTACAGAC 1187  
17 rArgAlaAspTyrLeulleLysLeuLeuSerArgAspLeuAlaLysArg 34  
1188 TCGTCAGATTACCTCATTAATTACTGAATAAAGACCTTCAAGAAAGG 1237  
34 luAlaGlnArgLeuCysGlyAla 41  
1238 AAGCACAGAGACTTGTGGTGCA 1260

seq\_name: /SIDS1/qcdata/hold-qeneseq/qeneseq-emb1/NA1997.DAT:AAT42751

seq\_documentation block:

seq\_documentation\_block:  
ID AAT42751 standard: cDNA: 6608 bp.

XX	
XX	
AC	
NC	
XX	AAT42751;
XX	
DT	12-MAR-1997 (first entry)
XX	
XX	Chicken CHD-1A gene.
XX	
KW	Bird; sex determination; chromodomain-Helicase-DNA binding 1 Avian;
KW	CHD-w; chromodomain-Helicase-DNA binding on the W chromosome; ss.

alignment_scores:		
Quality:	187.00	Length: 41
Ratio:	4.561	Gaps: 0
Percent Similarity:	100.000	Percent Identity: 87.805

alignment\_block:





CC sex in birds and can be used to identify the sex of an embryo,  
 CC foetus etc. and to manipulate the sex of progeny.

SQ Sequence 153 BP; 56 A; 36 C; 31 G; 30 T; 0 other;

# alignment\_scores:

Quality: 172.00 Length: 51  
 Ratio: 4.195 Gaps: 1  
 Percent Similarity: 80.392 Percent Identity: 70.588

# alignment\_block:

US-08-973-363-6 x AAT42758 ..

Align seg 1/1 to: AAT42758 from: 1 to: 153

```

1 IleLeuProAspAspProasp..... 7
1 ATTTTACCTGATGATCCAGATAAGAAACCCAGGCTAAGCAGTTACAGAC 50
8 .LysLysProGlnAlaLysGlnLeuGlnThrArgAlaAspTyrLeuIleL 24
|||||..... 100
51 CAAGAAACCCAGGCTAAGCAGTTACAGACCCGCTGCAGATTACCTCATTA 100
24 ysLeuLeuSerArgAspLeuAlaLysArgGluAlaGlnArgLeuCysGly 40
|||||..... 150
101 AATTACTGAATAAAGACCTTGCAAGAAAGGACACAGACAGCTTGCTGGT 150
41 Ala 41
|||
151 GCA 153

```

seq\_name: /SIDSl/gcgdata/hold-geneseq/geneseq-n-emb1/NA1997.DAT:AAT42759

# seq\_documentation\_block:

ID AAT42759 standard; DNA; 153 BP.

XX AC AAT42759;

XX DT 12-MAR-1997 (first entry)

XX DE Great tit CHD-W gene fragment.

XX KW Bird; sex determination; chromodomain-Helicase-DNA binding 1;

XX KW CHD-1A; CHD-W; W chromosome; ss.

XX OS Parus major.

XX FH Key Location/Qualifiers

FT misc\_difference 52...81

FT /\*tag= a

FT /note= "bases 52-81 are a repeat of bases 22-51 and are ignored in the translated amino acid sequence given in Fig 3"

XX PN WO9639505-A1.

XX PD 12-DEC-1996.

XX PF 05-JUN-1996; 96WO-GB01341.

XX PR 06-JUN-1995; 95GB-0011439.

XX PA (ISIS-) ISIS INNOVATION LTD.

XX PI Griffiths R, Tiwari B;

XX DR WPI; 1997-043127/04.

XX DR P-PSDB; AAW08149.

XX PT Avian chromodomain-helicase-DNA binding genes determine sex in

XX birds - used for sex determ. and to control sex of progeny

XX

PS Claim 8; Fig 3; 76pp; English.

CC Bases 3855-3977 (AAT42756) of the mouse CHD-1 gene show homology  
 CC to portions of the chicken CHD-1A (A = Avian) gene (AAT42757),  
 CC chicken CHD-W (W refers to the W chromosome) gene (AAT42758) and  
 CC and the great tit CHD-W gene (AAT42759). Translated amino acid  
 CC sequences of this region are provided in AAW08146-49. The CHD-1A  
 CC (see also AAT42751) and CHD-W (see also AAT42754-55) genes determine  
 CC sex in birds and can be used to identify the sex of an embryo,  
 CC foetus etc. and to manipulate the sex of progeny.

SQ Sequence 153 BP; 60 A; 33 C; 31 G; 29 T; 0 other;

# alignment\_scores:

Quality: 167.00 Length: 51  
 Ratio: 4.175 Gaps: 1  
 Percent Similarity: 78.431 Percent Identity: 68.627

# alignment\_block:

US-08-973-363-6 x AAT42759 ..

Align seg 1/1 to: AAT42759 from: 1 to: 153

```

1 IleLeuProAspAspProasp..... 7
1 ATTTTACCTGATGATCCAGATAAGAAACCCAGGCAAGCAGTTGCAGAC 50
8 .LysLysProGlnAlaLysGlnLeuGlnThrArgAlaAspTyrLeuIleL 24
|||||..... 100
51 CAAGAAACCCAGGCAAGCAGTTGCAGACCCGCTGCAGATTACCTCATTA 100
24 ysLeuLeuSerArgAspLeuAlaLysArgGluAlaGlnArgLeuCysGly 40
|||||..... 150
101 AATTACTGAATAAAGACCTTGCAAGAAAGAGAGTGCAGAGACTTACTGGT 150
41 Ala 41
|||
151 GCA 153

```

seq\_name: /SIDSl/gcgdata/hold-geneseq/geneseq-n-emb1/NA1998.DAT:AAV59280

# seq\_documentation\_block:

ID AAV59280 standard; cDNA; 1311 BP.

XX AC AAV59280;

XX DT 14-DEC-1998 (first entry)

XX DE Altered telomere repeat binding factor 1 gene.

XX KW ds; human; telomere repeat binding factor; A-TRF; dimerisation domain;

XX KW telomere; ageing; ataxia telangeictasia; Down's syndrome; tumour; viral.

XX OS Homo sapiens.

XX OS Synthetic.

XX FH Key Location/Qualifiers

FT CDS 1..1311

FT /\*tag= a

FT /product= "A-TRF"

XX PN WO9836066-A1.

XX PD 20-AUG-1998.

XX PF 13-FEB-1998; 98WO-US02765.

XX PR 04-FEB-1998; 98US-0018628.

XX PR 13-FEB-1997; 97US-0800264.

XX PA (UYRQ ) UNIV ROCKEFELLER.

XX

PI Bianchi A, De Lange T, Van Steensel B;  
 XX WPI; 1998-480769/41.  
 DR P-PSDB; AAV59280.  
 XX  
 PT Nucleic acid encoding altered telomere repeat binding protein and  
 related vectors - transformants, hetero-dimers and antibodies, used  
 to inhibit shortening of telomerases caused by ageing or disease,  
 PT also used to extend life of cells in culture  
 XX  
 PS Claim 14; Page 110-111; 163pp; English.  
 XX  
 CC The altered vertebrate telomere repeat binding protein (A-TRF) has a  
 CC telomere repeat binding factor (TRF) dimerisation domain, and forms a  
 CC hetero-dimer with TRF, preventing it from binding to the specified repeat  
 CC sequence. A-TRF, optionally expressed by gene therapy, is used to  
 CC inhibit shortening of telomeres associated with ageing (for cosmetic  
 CC purposes) and disease, e.g. ataxia telangiectasia, Down's syndrome,  
 CC atrophy of the skin, age-related macular degeneration, atherosclerosis,  
 CC tumours and viral (including human immune deficiency virus) infection.  
 CC Cells expressing A-TRF also have an increased life span *in vitro*, e.g.  
 CC for expression of recombinant proteins or where intended for subsequent  
 CC transplant or for testing, eliminating the need for transformation.  
 XX  
 SQ Sequence 1311 BP; 441 A; 257 C; 316 G; 297 T; 0 other;

alignment\_scores:  
 Quality: 129.00 Length: 25  
 Ratio: 5.160 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000  
 alignment\_block:  
 US-08-973-363-6 x AAV59280 ..  
 Align seg 1/1 to: AAV59280 from: 1 to: 1311  
 1 ILELEUPROASPSPROCAAPLYSLYSYPCGLNAlALYSGLNLEUlnTh 17  
 1237 ATTCTTCAGATGATCTCTGATAAAACACACACAAACAGTTACAGAC 1286  
 17 RAQGAlAAsPTVLEUILYSLEU 25  
 1287 CCGTGCAGACTACTCACTAICAACTA 1311

seq\_name: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAK88882  
 seq\_documentation\_block:  
 ID AAK8882 standard; cDNA; 421 BP.  
 XX  
 AC AAK88882;  
 XX  
 DT 05-NOV-2001 (first entry)  
 XX  
 DE Human digestive system antigen coding sequence SEQ ID NO: 1198.  
 XX  
 KW Human; digestive system antigen; gene therapy; cancer; appendicitis;  
 KW ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;  
 KW digestive system disorder; Meckel's diverticulum; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200155314-A2.  
 XX  
 PD 02-AUG-2001.  
 XX  
 PF 17-JAN-2001; 2001WO-US01324.  
 XX  
 PR 31-JAN-2000; 2000US-0179065.  
 PR 04-FEB-2000; 2000US-0180628.  
 PR 24-FEB-2000; 2000US-0184664.  
 PR 02-MAR-2000; 2000US-0186350.  
 PR 16-MAR-2000; 2000US-0189874.  
 PR

PR 17-MAR-2000; 2000US-0190076.  
 PR 18-APR-2000; 2000US-0198123.  
 PR 19-MAY-2000; 2000US-0205515.  
 PR 27-JUN-2000; 2000US-0209467.  
 PR 28-JUN-2000; 2000US-0214886.  
 PR 30-JUN-2000; 2000US-0215135.  
 PR 07-JUL-2000; 2000US-0216647.  
 PR 07-JUL-2000; 2000US-0216880.  
 PR 11-JUL-2000; 2000US-0217487.  
 PR 11-JUL-2000; 2000US-0217496.  
 PR 14-JUL-2000; 2000US-0218290.  
 PR 26-JUL-2000; 2000US-0220963.  
 PR 26-JUL-2000; 2000US-0220964.  
 PR 14-AUG-2000; 2000US-0224518.  
 PR 14-AUG-2000; 2000US-0224519.  
 PR 14-AUG-2000; 2000US-0225213.  
 PR 14-AUG-2000; 2000US-0225214.  
 PR 14-AUG-2000; 2000US-0225266.  
 PR 14-AUG-2000; 2000US-0225267.  
 PR 14-AUG-2000; 2000US-0225268.  
 PR 14-AUG-2000; 2000US-0225270.  
 PR 14-AUG-2000; 2000US-0225347.  
 PR 14-AUG-2000; 2000US-0225757.  
 PR 14-AUG-2000; 2000US-0225758.  
 PR 14-AUG-2000; 2000US-0225759.  
 PR 18-AUG-2000; 2000US-0226279.  
 PR 22-AUG-2000; 2000US-0226681.  
 PR 22-AUG-2000; 2000US-0226868.  
 PR 23-AUG-2000; 2000US-0227182.  
 PR 30-AUG-2000; 2000US-0227009.  
 PR 01-SEP-2000; 2000US-0228924.  
 PR 01-SEP-2000; 2000US-0229287.  
 PR 01-SEP-2000; 2000US-0229343.  
 PR 01-SEP-2000; 2000US-0229344.  
 PR 01-SEP-2000; 2000US-0229345.  
 PR 05-SEP-2000; 2000US-0229509.  
 PR 05-SEP-2000; 2000US-0229513.  
 PR 06-SEP-2000; 2000US-0230437.  
 PR 06-SEP-2000; 2000US-0230438.  
 PR 08-SEP-2000; 2000US-0231342.  
 PR 08-SEP-2000; 2000US-0231343.  
 PR 08-SEP-2000; 2000US-0231344.  
 PR 08-SEP-2000; 2000US-0231414.  
 PR 08-SEP-2000; 2000US-0232080.  
 PR 08-SEP-2000; 2000US-0232081.  
 PR 12-SEP-2000; 2000US-0231968.  
 PR 14-SEP-2000; 2000US-0232397.  
 PR 14-SEP-2000; 2000US-0232398.  
 PR 14-SEP-2000; 2000US-0232399.  
 PR 14-SEP-2000; 2000US-0232400.  
 PR 14-SEP-2000; 2000US-0232401.  
 PR 14-SEP-2000; 2000US-0233063.  
 PR 14-SEP-2000; 2000US-0233064.  
 PR 14-SEP-2000; 2000US-0233065.  
 PR 21-SEP-2000; 2000US-0234223.  
 PR 21-SEP-2000; 2000US-0234274.  
 PR 25-SEP-2000; 2000US-0234597.  
 PR 25-SEP-2000; 2000US-0234998.  
 PR 26-SEP-2000; 2000US-0235484.  
 PR 27-SEP-2000; 2000US-0235834.  
 PR 27-SEP-2000; 2000US-0235836.  
 PR 29-SEP-2000; 2000US-0236327.  
 PR 29-SEP-2000; 2000US-0236367.  
 PR 29-SEP-2000; 2000US-0236368.  
 PR 29-SEP-2000; 2000US-0236369.  
 PR 29-SEP-2000; 2000US-0236370.  
 PR 02-OCT-2000; 2000US-0236802.  
 PR 02-OCT-2000; 2000US-0237037.  
 PR 02-OCT-2000; 2000US-0237038.  
 PR 02-OCT-2000; 2000US-0237039.  
 PR 02-OCT-2000; 2000US-0237040.  
 PR 13-OCT-2000; 2000US-0239935.



PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226686.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241223.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249246.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251858.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Rosen CA, Barash SC, Ruben SM;  
XX WPI; 2001-457727/49.  
DR P-PSDB; AAM38625.  
XX  
XX Isolated polypeptide for treating, preventing and/or prognosing  
PT disorders related to the colon and rectum including colorectal cancers  
PT and also for testing and detection e.g. diagnosis.  
XX  
XX Claim 1; SEQ ID NO: 67; 522pp + Sequence Listing; English.  
PS  
XX The present invention provides the protein and coding sequences of a  
CC number of colorectal cancer antigens. These are shown in  
CC AAI57547-AAI57619 and AAM38569-AAM38641. These can be used in the  
CC diagnosis, prevention and treatment of cancer of the colon and/or rectum.  
CC The present sequence is a colorectal cancer antigen coding sequence of  
CC the invention.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 421 BP; 148 A; 78 C; 127 G; 61 T; 7 other;  
alignment\_scores: Length: 35  
Quality: 118.00  
Ratio: 4.069  
Gaps: 0





CC a26g1, and encode type I polyketide synthases.  
XX  
SQ Sequence 5292 BP; 905 A; 1715 C; 1725 G; 947 T; 0 other;

alignment\_scores:  
Quality: 58.00 Length: 40  
Ratio: 2.231 Gaps: 2  
Percent Similarity: 65.000 Percent Identity: 40.000

alignment\_block:  
US-08-973-363-6 x AAF90036/rev ..

Align seg 1/1 to reverse of: AAF90036 from: 1 to: 5292

5 AspProAspLysLysProGlnAlaLysGln...LeuGlnThrArgAlaAs 20  
||| ||| ::|||::: ::||| ::| ::| |||||::  
4695 GACAGGACGTGGCGCGGAATCTGCTCAATGCTTTGTAGAGCGCGCA 4646

20 pTyrLeuIleLysLeuSerArgAspLeuAlaLysArgGlu..... 34  
: |||:::|||| ::||| |||::: ::|

4645 ATCCCTGTTGGGTTGAATAACCGAGATGCGCGCGCGCCAGACGGTCT 4596

35 ..AlaGlnArgLeuCysGly 40  
||| ||||| |||||

4595 CCGCGGTCGCTGTCGGC 4576

seq\_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAF90033

seq\_documentation\_block:

ID AAF90033 standard; DNA; 34071 BP.

XX AC AAF90033;

XX DT 06-AUG-2001 (first entry)

XX DE Nucleotide sequence of cosmid a26g1 (coding strand).

XX KW Metabolic pathway operon; polyketide; polyketide antibiotic; ss.

XX OS Synthetic.

XX PN WO200140497-A2.

XX PD 07-JUN-2001.

XX PF 27-NOV-2000; 2000WO-FR03311.

XX PR 29-NOV-1999; 99FR-0015032.

XX PR 07-JUN-2000; 2000US-0209800.

XX PA (AVET ) AVENTIS PHARMA SA.

XX PI Jeannin P, Pernodet J, Guerineau M, Simonet P, Courtois S;  
PI Cappellano C, Francou F, Raynal A, Ball M, Sezonov G, Tuphile K;  
PI Frostegard A;

XX DR WPI; 2001-374849/39.

XX Collection of nucleic acids from environmental samples, useful for  
PT identifying e.g. genes encoding polyketide synthases and derived  
PT antibiotics -

XX Example 14; Page 289-300; 356pp; French.

XX The specification describes a method for the preparation of a collection  
CC of nucleic acids from organisms in a soil sample. The method comprises  
CC milling a dried sample to produce microparticles; suspending these in  
CC liquid buffer; extraction of nucleic acids from the microparticle;  
CC passing nucleic acid-containing solution through a molecular sieve;  
CC passing nucleic acid-enriched fractions through an anion exchange  
CC chromatography material; and recovering fractions containing purified  
CC nucleic acids. The nucleic acids are sources for sequences that encode

CC either operons involved in a metabolic pathway (specifically polyketide  
CC synthesis) or polypeptides, particularly for production of therapeutic  
CC or agricultural compounds, especially polyketide antibiotics. The present  
CC sequence represents cosmid a26g1 coding strand), which encodes different  
CC polyketide synthases.  
XX  
SQ Sequence 34071 BP; 5791 A; 10858 C; 11089 G; 6333 T; 0 other;

alignment\_scores:  
Quality: 58.00 Length: 40  
Ratio: 2.231 Gaps: 2  
Percent Similarity: 65.000 Percent Identity: 40.000

alignment\_block:

US-08-973-363-6 x AAF90033/rev ..

Align seg 1/1 to reverse of: AAF90033 from: 1 to: 34071

5 AspProAspLysLysProGlnAlaLysGln...LeuGlnThrArgAlaAs 20  
||| ||| ::|||::: ::||| ::| ::| |||||::  
17630 GACAGGACGTGGCGCGGAATCTGCTCAATGCTTTGTAGAGCGCGCA 17581

20 pTyrLeuIleLysLeuSerArgAspLeuAlaLysArgGlu..... 34  
: |||:::|||| ::||| |||::: ::|

17580 ATCCCTGTTGGGTTGAATAACCGAGATGCGCGCGCGCCAGACGGTCT 17531

35 ..AlaGlnArgLeuCysGly 40  
||| ||||| |||||

17530 CCGCGGTCGCTGTCGGC 17511

seq\_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAF90032

seq\_documentation\_block:

ID AAF90032 standard; DNA; 42717 BP.

XX AC AAF90032;

XX DT 06-AUG-2001 (first entry)

XX DE Nucleotide sequence of cosmid a26g1 (non-coding strand).

XX KW Metabolic pathway operon; polyketide; polyketide antibiotic; ss.

XX OS Synthetic.

XX PN WO200140497-A2.

XX PD 07-JUN-2001.

XX PF 27-NOV-2000; 2000WO-FR03311.

XX PR 29-NOV-1999; 99FR-0015032.

XX PR 07-JUN-2000; 2000US-0209800.

XX PA (AVET ) AVENTIS PHARMA SA.

XX PI Jeannin P, Pernodet J, Guerineau M, Simonet P, Courtois S;  
PI Cappellano C, Francou F, Raynal A, Ball M, Sezonov G, Tuphile K;  
PI Frostegard A;

XX DR WPI; 2001-374849/39.

XX Collection of nucleic acids from environmental samples, useful for  
PT identifying e.g. genes encoding polyketide synthases and derived  
PT antibiotics -

XX Example 14; Page 274-288; 356pp; French.

XX The specification describes a method for the preparation of a collection  
CC of nucleic acids from organisms in a soil sample. The method comprises  
CC milling a dried sample to produce microparticles; suspending these in  
CC liquid buffer; extraction of nucleic acids from the microparticle;



passing nucleic acid-containing solution through a molecular sieve; passing nucleic acid-enriched fractions through an anion exchange chromatography material; and recovering fractions containing purified nucleic acids. The nucleic acids are sources for sequences that encode either operons involved in a metabolic pathway (specifically polyketide synthesis) or polypeptides, particularly for production of therapeutic or agricultural compounds, especially polyketide antibiotics. The present sequence represents cosmid a2691 (non-coding strand). The sense strand encodes different polyketide synthases.

Sequence 42717 BP; 8230 A; 13520 C; 13184 G; 7782 T; 1 other;

```
alignment_scores:
  Quality: 58.00      Length: 40
  Ratio: 2.231       Gaps: 2
  Percent Similarity: 65.000    Percent Identity: 40.000
```

```
alignment_block:
US-08-973-363-6 x AAF90032
```

Align seq 1/1 to: AAF90032 from: 1 to: 42717

[illegible]

35 ..AlaGlnArqLeuCysGly 40

16590 CCGCGGTTCGCCCTGTGCGGC 16609

BEST AVAILABLE COPY

THIS PAGE BLANK (USPTO)